MICROBIOLOGY AND MOLECULAR BIOLOGY REVIEWS, June 2008, p. 228–248 Vol. 72, No. 2
1092-2172/08/$08.00+0 doi:10.1128/MMBR.00041-07

Evolutionary Persistence of the Molybdopyranopterin-Containing Sulfite Oxidase Protein Fold

Gregory J. Workun, Kamila Moquin, Richard A. Rothery, and Joel H. Weiner*

Membrane Protein Research Group, Department of Biochemistry, University of Alberta, 474 Medical Sciences Building, Edmonton, Alberta T6G 2H7, Canada

INTRODUCTION

Enzymes containing the metal molybdenum or tungsten at their active sites are essential for nearly all living organisms and catalyze primarily oxo transfer reactions for the metabolism or catabolism of nitrogen, sulfur, and carbon compounds (20, 34–37, 46, 57, 58). For all examples other than those belonging to the nitrogenase class (18, 69), the Mo or W is coordinated by a pyranopterin (PPT) cofactor (35, 37). This coordination is supplemented by an amino acid side chain such as those of Cys, Sec, Ser, or Asp residues and/or by the coordination of single oxygen or sulfur atoms in the form of oxo or sulfido groups. Molybdoenzymes have been assigned to four classes. The first class is the archaeal aldehyde oxidoreductase class that contains a bis-PPT cofactor that coordinates a W atom at its active site (10, 66). This class lacks a protein W ligand. The second class is the xanthine dehydrogenase class, examples of which include the eukaryotic xanthine dehydrogenase class and *Desulfovibrio gigas* aldehyde oxidoreductase (36). This class also lacks a protein-Mo ligand. The third class

* Corresponding author. Mailing address: Membrane Protein Research Group, Department of Biochemistry, University of Alberta, 474 Medical Sciences Building, Edmonton, Alberta T6G 2H7, Canada. Phone: (780) 492-2761. Fax: (780) 492-0886. E-mail: joel.weiner@ualberta.ca.

is the prokaryotic dimethyl sulfoxide reductase class that contains a Mo-bis(pyranopterin guanine dinucleotide) (Mo-bisPGD) cofactor (5, 39, 75, 80, 81) with the full range of protein-Mo ligands described above. The fourth class is the sulfite oxidase (SUOX) class that contains a Mo-PPT cofactor and a Cys protein-Mo ligand (46, 47). The plant-type assimilatory nitrate reductase (NIA) also belongs to the SUOX class (8, 22) and functions in nitrate assimilation (56). Animal SUOX enzymes funnel electrons derived from sulfite oxidation to cytochrome *c* within the intermembrane space of the mitochondrion. Plant SUOX enzymes are located in the peroxisome and transfer sulfite-derived electrons directly to oxygen, generating hydrogen peroxide (30, 31). There are also a large number of archaeal and bacterial enzymes of currently unknown function that are predicted to contain an SUOX fold.

REACTIONS CATALYZED BY SUOX FOLD ENZYMES

The catalytic mechanisms of the SUOX enzymes have been reviewed extensively (20, 30, 31, 46) and will not be dealt with in detail herein. Animal SUOX catalyzes the following reaction in the mitochondrial intermembrane space with cytochrome *c* as the electron acceptor:

$$SO_3^{2-} + H_2O + 2c_{ox} \rightarrow SO_4^{2-} + 2c_{red} + 2H^+.$$

Plant SUOX catalyzes a similar reaction but with oxygen as an electron acceptor (30–32):

$$SO_3^{2-} + O_2 + H_2O \rightarrow SO_4^{2-} + H_2O_2.$$

The reaction catalyzed by plant NIA is as follows:

$$NO_3^- + NADH + 2H^+ \rightarrow NO_2^- + NAD^+ + H_2O.$$

In each case, the oxidation (ox) of sulfite or the reduction (red) of nitrate occurs at the Mo atom of a Mo-PPT cofactor (20, 22, 34, 37, 46). In fully oxidized SUOX and NIA, Mo(VI) coordination is provided by a pair of pterin-derived dithiolene sulfurs, the sulfur of a Cys residue, and two oxo groups (14, 20, 22). A single electron reduction generates the Mo(V) form, which retains the dithiolene coordination and has one oxo group and one hydroxyl or water ligand. The fully reduced Mo(IV) form has coordination very similar to that of the Mo(V) form (14, 33). The proposed SUOX reaction cycle envisions bond formation between the sulfite lone pair and a Mo(VI) oxo group and oxidative oxo transfer to generate the product sulfate and the Mo(IV) form of the enzyme. Thus, oxo transfer occurs from the Mo atom to the substrate. The NIA reaction would proceed in the reverse direction, with oxo transfer occurring from the substrate nitrate to the Mo atom.

Of the bacterial enzymes predicted to share the SUOX fold, the best characterized is SUOX from *Starkeya novella*. This enzyme has been described as being a "sulfite dehydrogenase" in the literature (11). It is a heterodimer comprising a SUOX fold protein (SorA) and a cytochrome *c* subunit (SorB) (41, 42). SorAB is periplasmically localized and is predicted to transfer electrons from sulfite oxidation to the membrane-bound respiratory chain via at least one soluble cytochrome *c* (43, 92).

Another partially characterized bacterial SUOX fold enzyme system is encoded by the *Escherichia coli yedYZ* operon. YedY is a periplasmically localized SUOX fold protein that appears to interact transiently with membrane-bound cytochrome *b* (YedZ) (7, 53). The physiological substrate for YedY is currently unknown, although it demonstrates reduction of a range of S- and N-oxides. Data on two further bacterial SUOX enzymes are also available. These enzymes are the cytoplasmically located SUOX from *Deinococcus radiodurans* (12) and the periplasmically localized SUOX from *Thermus thermophilus* (13). Neither of these enzymes appears to have subsidiary cytochrome subunits or domains, although the latter enzyme is predicted to be periplasmically localized and to be redox coupled to the cytoplasmic membrane via cytochrome *c*.

SUOX FOLD ENZYMES IN HUMAN DISEASE AND GEOCHEMICAL CYCLES

Human SUOX deficiency is typically inherited as a recessive autosomal trait for which there is no known therapy and typically results in death in infancy (38). Its severity is due to the critical role of SUOX in degrading Cys and Met amino acids, and deficiency results in the excretion of large amounts of *S*-sulfo-L-cysteine, sulfite, and thiosulfate in the urine (65). The intracellular accumulation of sulfite is thought to result in sulfitolysis of disulfide bonds that results in compromised protein stability (20). Point mutations linked to SUOX deficiency include the following: R160Q, A208D, S370Y, and G473D (47). The R160Q mutation increases the K_m for sulfite and decreases the k_{cat}, resulting in a 1,000-fold decrease in catalytic efficiency (23). Structural analysis of chicken liver SUOX reveals the importance of three Arg residues (47). R138, R190, and R450 contribute to a positively charged binding pocket, which stabilizes substrate/product binding, and these residues will be referred to herein as the Arg triad (see Active-Site Sequence Conservation Suggests a Simplified Classification of SUOX Fold Enzymes). The effects of an R160Q mutant in humans, corresponding to R138Q in chicken SUOX, have recently been studied by X-ray absorption spectroscopy (15) and reveal an increase in coordination number for the Mo, from 5 to 6. In the case of the R138Q mutant in chicken SUOX, X-ray crystallography revealed that the side chain nitrogen of the Gln appears to be within the coordination sphere of the Mo (44).

Plant SUOX functions in sulfite detoxification and has been implicated in the adaptation to elevated sulfur dioxide levels ("acid rain") (30, 32). Atmospheric sulfur dioxide is converted to sulfite as follows:

$$SO_2 + H_2O \rightarrow [SO_2 \cdot H_2O] \rightarrow HSO_3^- + H^+ \leftrightarrow SO_3^{2-} + 2H^+.$$

SUOX is thus of importance to biosphere sulfur cycling and adaptation to industrial pollution. Sulfate is the major anion within the xylem system of plants and is absorbed by the root system for biosynthetic incorporation into Cys/Met residues and also into other compounds such as glucosinolates (6).

One novel aspect of plant SUOX is its reactivity toward oxygen, which results in the generation of potentially toxic H_2O_2 during enzyme turnover (20, 67). This problem is alleviated by the cellular localization of plant SUOX to the peroxisome organelle, which contains relatively large concentrations of peroxidase/catalase enzymes (52).

Plant NIA enzymes play a crucial role in nitrogen uptake and assimilation into plant proteins and play a role in the

global nitrogen cycle (56, 72). They play a role complementary to that of the distinct nitrate reductases found in bacterial systems, whose activities are dependent on the Mo-bisPGD cofactor (73, 75).

EMERGING STRUCTURES REVEAL MULTIDOMAIN ARCHITECTURES OF SUOX FOLD ENZYMES

The importance of SUOX fold enzymes in human disease, plant sulfite detoxification, agriculture, adaptation to pollution, and global geochemical cycles has prompted efforts to obtain atomic-resolution structural data by protein crystallography and X-ray absorption spectroscopy methods. Structures for animal, plant, and bacterial SUOX enzymes, as well as the Mo-PPT binding SUOX fold fragment of a plant-type NIA enzyme, have been reported (14, 15, 33, 43, 44, 47, 82). Analyses of these structures reveal that animal SUOX enzymes comprise a heme b_5 domain, a Mo-PPT binding SUOX fold, and a dimerization domain, with this order being reflected in the primary sequence of the protein. The available plant-type NIA structure constitutes a fragment of the complete enzyme and includes the SUOX domain and the dimerization domain (22). Plant SUOX comprises a SUOX fold and a dimerization domain, and it lacks a heme-containing domain or accessory subunit and is localized to the peroxisome organelle (30, 31, 57). The periplasmic SUOX from *S. novella* has a Mo-PPT binding catalytic subunit (SorA) comprised of an SUOX fold and a dimerization domain. SorA forms a heterodimer with an additional subunit, SorB, which coordinates a *c*-type heme (41, 43).

Genome and protein structure database mining has identified a range of bacterial enzymes sharing sequence and inferred structural similarity to the well-characterized members of the SUOX fold-containing family. In the case of *E. coli*, YedY, a SUOX fold protein, has been identified, and its structure has been solved by protein crystallography (7, 53). Intriguingly, it is comprised of only the Mo-PPT binding SUOX fold itself. The redox partner of YedY is YedZ, a hydrophobic monoheme cytochrome *b* (midpoint potential [E_m] value -8 mV) (7) with six predicted transmembrane (TM) segments (17) (see Differential Redox Partners) that presumably couples YedY to the versatile redox sink of the membrane-intrinsic quinone pool. It was demonstrated using PSI-BLAST searches, which identify homologs with low similarity to bait sequences, that a plasma membrane antigen that is overexpressed in prostate cancers is related to YedZ (77, 89).

Another variation of the multidomain architecture observed in the SUOX fold enzyme is provided by the plant-type NIA enzymes that are also represented in yeasts such as *Pichia angusta* (22). These retain the SUOX fold and the dimerization domain, but the heme b_5 domain follows the dimerization domain in the primary sequence. Also, it contains an additional flavin adenine dinucleotide (FAD)-containing NADH binding domain at its C terminus, which, in the case of *Zea mays*, has had its structure solved by protein crystallography (54).

Our interest in the structure and function of the SUOX fold proteins stems from our discovery of the YedYZ system of *E. coli* (7, 53). The physiologically relevant reaction catalyzed at the Mo-PPT of this enzyme remains unresolved. As discussed below (see SUOX Fold Protein Cladistics), YedY and its predicted redox partner, YedZ, are highly conserved across a broad range of primarily proteobacterial species, so it likely catalyzes a conserved interconversion within their metabolomes.

SECONDARY STRUCTURE MATCHING SEARCHES REVEAL SIMILARITIES BETWEEN PROTEINS OF LOW SEQUENCE SIMILARITY

The availability of several hundred genome sequences presents an excellent opportunity to analyze SUOX fold proteins in all domains of life. As of August 2007, the complete sequences of 46 archaeal, 523 bacterial, and 65 eukaryotic genomes were available (listed on the Genomes Online Database at http://www.genomesonline.org). This is reflected in the total number of protein sequences available, which exceeds 4.4 million as of August 2007 (listed on the European Bioinformatics Institute [EBI] server at http://www.ebi.ac.uk/swissprot/sptr_stats/index.html). Remarkable strides have also been made in protein structure determinations, with approximately 47,000 structures being available as of August 2007 (listed on the Protein Data Bank [PDB] server at http://www.pdb.org). The sheer size of available sequence and structural bioinformatic data sets has catalyzed the emergence of a diverse array of analysis methods. One such approach is the use of secondary structure matching (SSM) algorithms to compare a protein structure of interest with either a set of protein structures or the entire PDB archive (49). We have recently used this technique to gain new insights into the complex iron-sulfur molybdoenzyme family (75).

In order to study the relationships between the SUOX fold proteins, we used the EBI SSM server (at www.ebi.ac.uk/msd-srv/ssm/ssmstart.html) (49) to search the entire PDB archive (at www.pdb.org). Because of its simplified domain composition, the structure of *E. coli* YedY (PDB accession number 1XDQ) (chain A, 2.6-Å resolution) was used as bait (53). Table 1 summarizes the results of this "three-dimensional BLAST search". YedY shows striking structural similarity to the following enzymes: plant SUOX (from *Arabidopsis thaliana* [PDB accession number 1OGP] [2.6-Å resolution]) (82), chicken SUOX (from *Gallus gallus* [accession number 1SOX] [1.9-Å resolution]) (47), NIA from *P. angusta* (accession number 2BII) (1.7-Å resolution) (22), and SUOX from *S. novella* (accession number 2BLF) (1.8-Å resolution) (41). Table 1 reveals that between 66 and 70% of maximum possible C-α positions overlap between the bait and hit structures.

Figure 1 illustrates the similarities and differences among the five structures. Each of the hits is significantly larger than the bait structure (YedY [PDB accession number 1XDQ]), with the latter comprising a Mo-PPT-coordinating catalytic core and being considered to be a prototypical SUOX fold (see Evolutionary Pathway of the SUOX Fold Proteins). The SUOX fold of the enzymes shown in Fig. 1 comprises a mixed α and β structure organized as two to three β-sheets and 9 to 12 α-helices (47, 53, 82). Each of the four hit proteins contains the distinctive seven-stranded β-barrel structure characteristic of an SUOX dimerization domain (47). SUOX from *G. gallus* also contains an N-terminal heme b_5 domain. SUOX from *S. novella* (SorAB) has the additional cytochrome *c* subunit (SorB) (11, 43, 92) that plays a role similar to that of the heme b_5 domain of *G. gallus* SUOX (41). Plant SUOX lacks an additional heme-containing subunit or domain (82). Finally, *P. angusta* NIA resembles the enzyme found in plants, which comprises a SUOX fold followed by a dimerization do-

TABLE 1. SSM results by using the structure of *E. coli* YedY (PDB accession number 1XDQ) as bait

PDB accession no.	Description	Resolution (Å)	RMSD[a] (Å)	No. of identified backbone C-α atoms	No. of overlapping C-α atoms between the two structures[b]	$ID_{C-\alpha}$[c]	No. of gaps[d]
1XDQ	*E. coli* YedY	2.6	0	262	262	100	0
1OGP	*Arabidopsis thaliana* SUOX	2.6	2.02	388	185	71	11
2BLF	*Starkeya novella* SUOX	1.8	1.85	373	173	66	11
2BII	*Pichia angusta* NIA	1.7	2.00	415	182	69	10
1SOX	*Gallus gallus* SUOX	1.9	2.03	463	181	69	14

[a] Deviation between PDB accession number 1XDQ and the indicated structure.
[b] Number of overlapping C-α atoms between the two structures with the quoted RMSD.
[c] $ID_{C-\alpha}$, C-α position identity expressed as a percentage of the maximum possible.
[d] Number of gaps represented in the sequences of the structural alignment.

main, a heme b_5 domain, and a FAD-containing NADH binding domain. The structure determined by Fischer et al. (22) represents only the N-terminal SUOX fold and the dimerization domain. A structure is available for the NADH binding domain from the NIA enzyme from *Z. mays* at a resolution of 2.5 Å (54). With the exception of YedY, each of the SUOX fold enzymes has a well-defined substrate for which a wealth of kinetic data has been collected.

SSM GENERATES AN IMPROVED SEQUENCE ALIGNMENT OF SUOX FOLD ENZYMES

The availability of five experimentally determined protein structures containing the SUOX fold presents an opportunity to investigate the use of SSM matching to generate an improved sequence alignment of these proteins. Using CLUSTALW (87, 88), we generated an alignment inconsistent with the structural

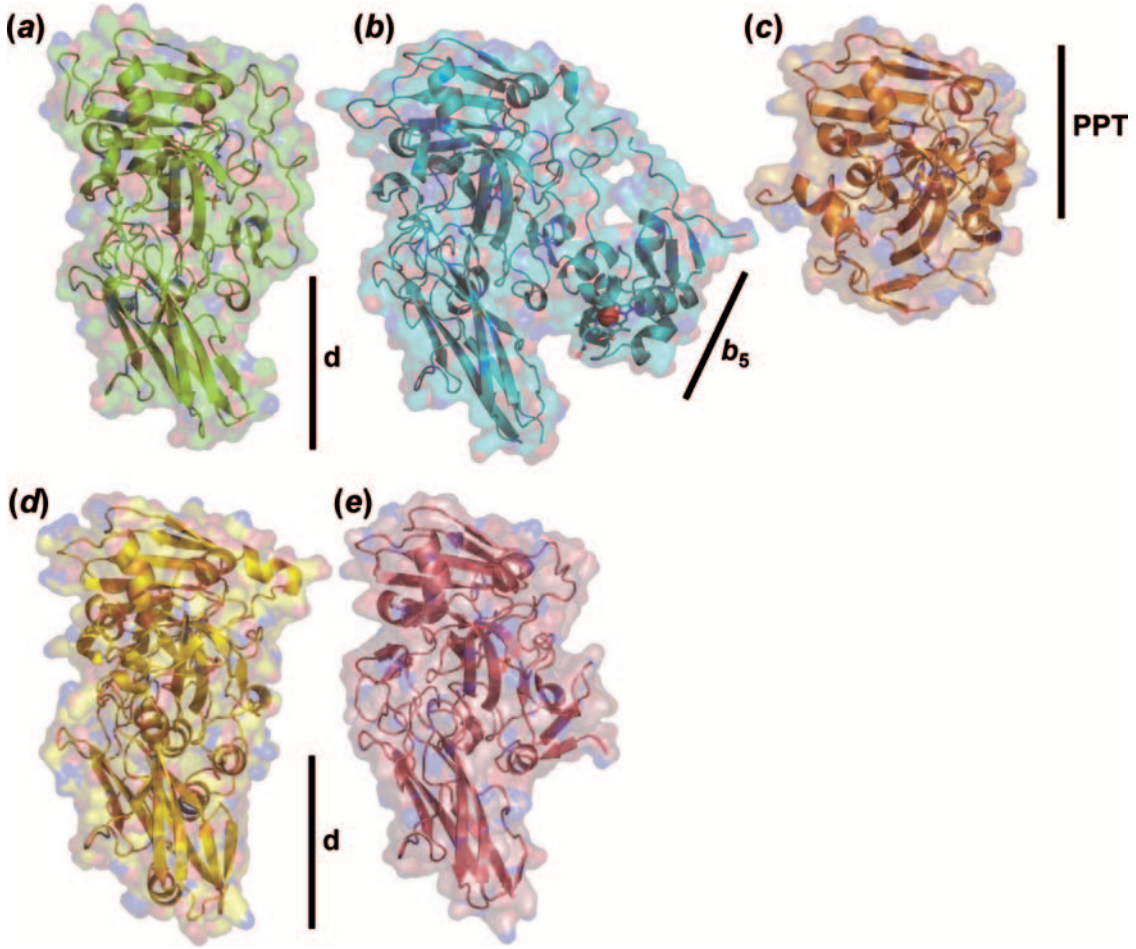

FIG. 1. Unique structures sharing the SUOX fold. (a) Plant SUOX (PDB accession number 1OGP). (b) Chicken SUOX (accession number 1SOX). (c) *E. coli* YedY (accession number 1XDQ). (d) *P. angusta* NIA (accession number 2BII). (e) *S. novella* SUOX (accession number 2BLF). The positions of the PPT binding, dimerization, and b_5-binding domains are labeled PPT, d, and b_5, respectively. The figure was created with the PYMOL molecular visualization package (DeLano Scientific LLC, Palo Alto, CA).

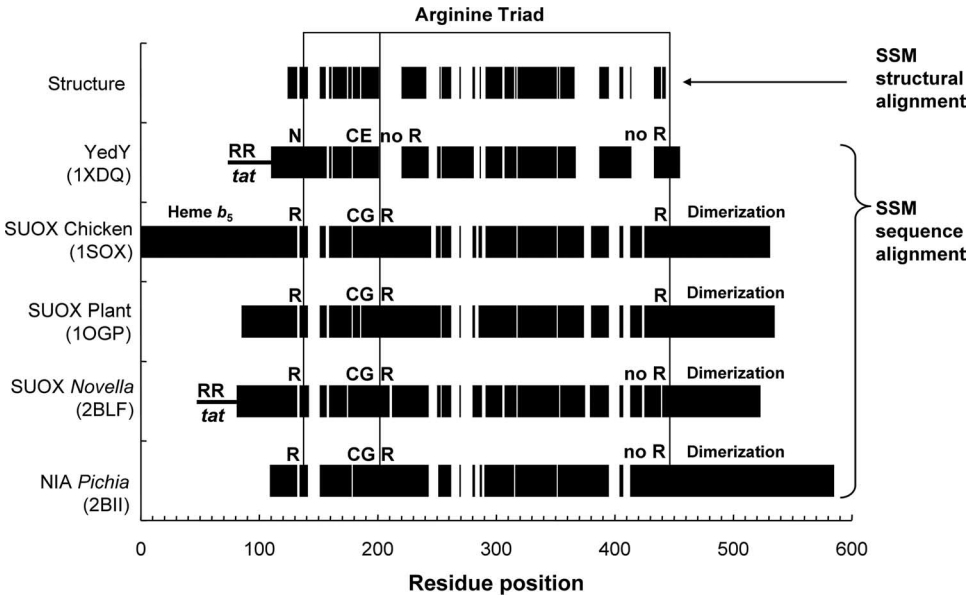

FIG. 2. Comparison of secondary structure alignment with sequence features in the five proteins sharing the SUOX fold. The blocks of the upper bar graph represent segments of sequences from the five structures that overlap over 181 C-α positions (out of a total possible of 262) with an RMSD of 1.82 Å in the multiple alignment. It can also be described as the sequence that constitutes the core of the SUOX fold. Gaps represent sequence segments that either do not align or are absent in one or more of the proteins. The lower five bar graphs represent the alignment of the five sequences based on the SSM alignment. In these cases, gaps indicate the absence of sequence. The presence of the heme b_5 domain at the N terminus of chicken SUOX (PDB accession number 1SOX) is indicated, as is the presence of the dimerization domain in all the structures except for that of YedY (accession number 1XDQ). YedY and *S. novella* also possess an N-terminal *tat* leader that is absent from the mature (crystallized) protein. See the text for a discussion of critical Arg and other residues within the five proteins.

overlap implicit in Fig. 1 (data not shown). Figure 2 shows a graphical alignment of structural conservation and sequence within the group of five structures that is based on SSM analyses. It also shows the approximate positions in the primary sequences of five critical active-site residues (see below), the heme b_5 domain, and the dimerization domain. Also shown is the N-terminal *tat* leader that directs YedY and other cofactor-containing enzymes to the periplasmic compartment of *E. coli* and other bacteria (1, 3, 90). Figure 3 presents the textual alignment resulting from the SSM approach. Within the multiple alignment, a total of 181 C-α positions (out of a total possible of 262, the number of C-α positions in the YedY structure [PDB accession number 1XDQ]) overlap with an overall root mean square deviation (RMSD) of 1.82 Å. A total of 15 residues are absolutely conserved within the five structures, including the Mo-coordinating Cys residue (YedY-C102, aligned to position 88 in Fig. 3), and the positions of these residues within the structure of YedY (accession number 1XDQ) are indicated in Fig. 4. Remarkably, except for the Cys protein-Mo ligand, only three of these conserved residues, corresponding to YedY-K207, YedY-R194, and YedY-G202, are within 4 Å of the Mo-PPT cofactor. In the case of the two inferred cationic side chains (K207 and R194), the side chains are within 3 Å of the phosphate oxygens of the phosphate group of the cofactor, presumably interacting electrostatically with it. As described below, none of these residues appear to be implicated in the active-site funnels of the SUOX fold-containing family. Two further residues appear to be close to the Mo-PPT shown in Fig. 4, Val-133 and Pro-198; however, this is an artifact of the projection used, with these residues being ~14 Å and ~11.9 Å away, respectively.

Given the complementary nature of the SSM and sequence-based bioinformatics methods, we compared the sequence identity data obtained from the SSM alignment (Fig. 3) with data obtained from pairwise sequence comparisons using the EMBOSS STRETCHER program (70). Figure 5 shows the relationship between pairwise RMSD values obtained from SSM analyses and equivalent pairwise sequence identities. It is apparent that sequences that are deemed quite dissimilar in conventional bioinformatic analyses, with sequence identities of ≤20%, can overlap, with strikingly low RMSD values of ≤2 Å. For example, *E. coli* YedY (PDB accession number 1XDQ) and *P. angusta* NIA (accession number 2BII) overlap, with an RMSD of 2 Å, but are only 11.5% identical in sequence. At the other extreme, plant (accession number 1OGP) and chicken (accession number 1SOX) SUOXs overlap, with an RMSD of 1.02 Å, and are 38.3% identical in sequence. However, when identity is defined as a percentage of possible C-α overlap, these two pairs exhibit 69% and 92% identities, respectively. Overall, these observations illustrate the power of the SSM methodology in exploring structural and sequence identities within protein structure families.

ACTIVE-SITE SEQUENCE CONSERVATION SUGGESTS A SIMPLIFIED CLASSIFICATION OF SUOX ENZYMES

Figure 6 shows a simplified view of the active-site region of structurally characterized members of the SUOX fold family.

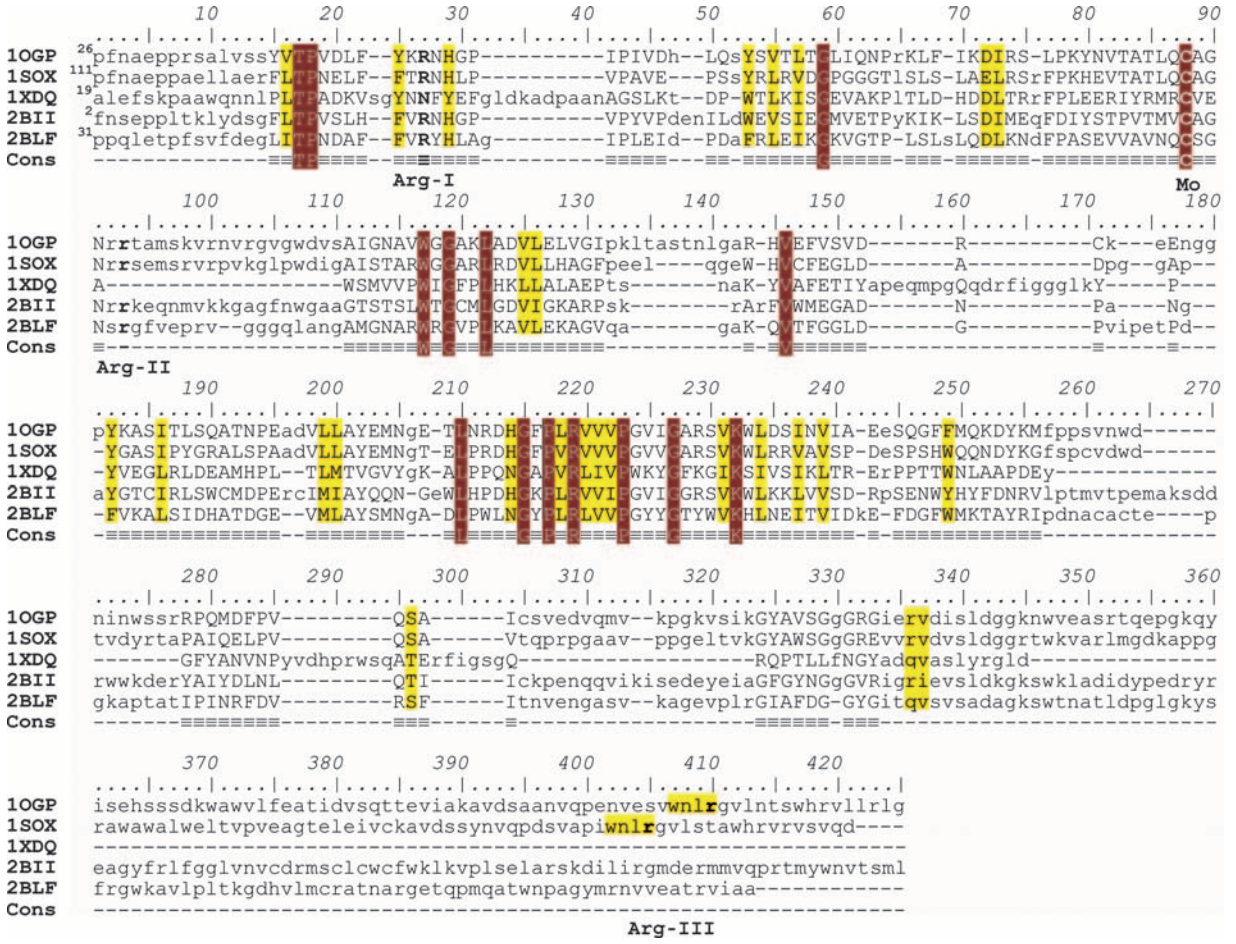

FIG. 3. Sequence alignment based on SSM matching analyses of proteins in the PDB sharing the SUOX fold. The following proteins were selected for analysis: plant SUOX (PDB accession number 1OGP) (82), chicken SUOX (accession number 1SOX) (47), *E. coli* YedY (accession number 1XDQ) (53), *Pichia* nitrate reductase (accession number 2BII) (22), and *S. novella* sulfite dehydrogenase (accession number 2BLF) (41). Capitalized single-letter abbreviations represent the 181 residues whose C-α positions overlap with an RMSD of 1.82 Å. Structural overlap is also indicated in the consensus sequence by use of the "≡" symbol. The 15 absolutely conserved residues are indicated. For the sake of brevity, the N-terminal 110 residues of chicken SUOX (accession number 1OGP) and the C-terminal 50 residues of *Pichia* NIA are not shown. The positions of the conserved substrate-binding Arg residues and the Cys protein-Mo ligand are indicated.

Each of the enzymes has a conserved Cys residue that provides the protein-Mo ligand, with the rest of the Mo coordination sphere being provided by the dithiolene sulfurs of the PPT and two oxo groups (14, 15). With the exception of YedY, each of the enzymes has at least two Arg residues in its active site, and these have been proposed to stabilize the negative charge of the substrate (sulfite and nitrate) or product (sulfate and nitrite) (21, 44, 47). An additional feature in terms of sequence conservation is the presence of a Gly residue following the Cys. A Glu appears at this position in YedY. This rather simplistic description of active-site residues enables the categorization of the SUOX fold enzymes based on active-site conservation and greatly simplifies the interpretation of the cladistics presented below (see SUOX Fold Protein Cladistics). Genuine SUOX enzymes can be coded as R C G R R: they retain the "arginine triad," have the conserved Cys as a protein-Mo ligand, and retain the active-site Gly. A critical non-active-site feature of enzymes containing the SUOX fold is the presence or absence of a *tat* leader (3, 90). YedY, for example, can be coded *tat* N C E | |, indicating that it has a *tat* leader in its immature form, an Asn at the first Arg triad position ("N"), a Cys protein-Mo ligand, a Glu ("E") at the position occupied by a Gly in many of the other SUOX fold proteins, and an absence of residues corresponding to the final two Arg residues of the Arg triad ("| |"). This coding will be discussed in more detail when the cladistics of the SUOX fold family of proteins are considered below.

SUOX FOLD PROTEIN CLADISTICS

Our observations on the structural conservation of the SUOX fold in organisms ranging from *E. coli* to *G. gallus* prompted us to investigate the full range of organisms in which it occurs. Because of the relatively limited number of SUOX fold structures from a broad taxonomic diversity of organisms, we relied upon a traditional sequence-based bioinformatics approach. The sequences of the proteins sharing significant structural similarity with *E. coli* YedY were each used as bait

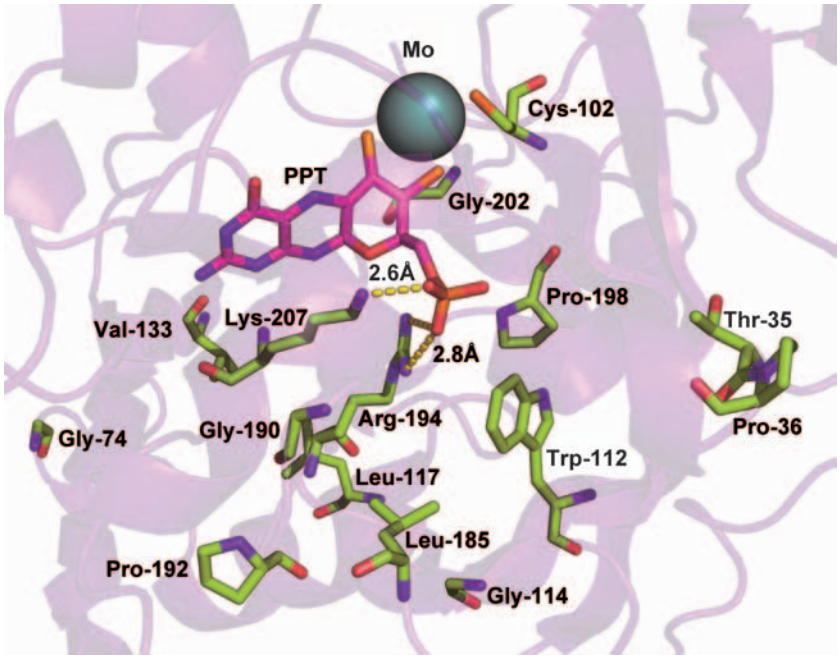

FIG. 4. Locations of the 15 conserved residues within the family of structurally characterized SUOX fold proteins. Numbering is done according to *E. coli* YedY (PDB accession number 1XDQ), and the conserved residues are displayed, with the backbone rendered in PYMOL in transparent cartoon format. Of the 15 conserved residues, only Cys102 (protein-Mo ligand), Gly202, Lys207, and Arg194 are within 4 Å of the Mo-PPT.

to search the UNIPROT sequence database using the BLASTP algorithm at the EBI Wu-BLAST2 server (www.ebi.ac.uk/blast2/; developed by W. R. Gish). Accession numbers from each of the searches were merged, and duplicate entries were eliminated. The EMBOSS SEQRET program (70) was used to retrieve the entire sequence set corresponding to the hits in the Wu-BLAST2 searches. This generated a sequence set of 559 sequences that were used to generate a CLUSTALW alignment (87, 88), and obvious outliers were removed by inspection. In order to generate a suitable data set for evolutionary analyses, the sequences were filtered pairwise to remove sequences with greater than 60% identity to any other sequence in the data set (using the EMBOSS STRETCHER

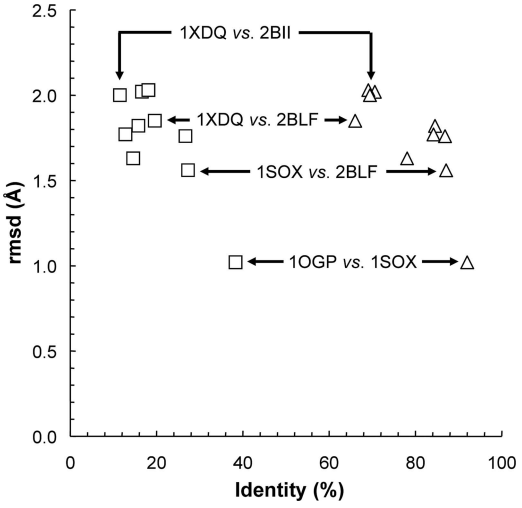

FIG. 5. Relationship between sequence identity and SSM RMSD values within the SUOX family. Data for the plot of RMSD versus sequence identities (squares) were obtained by doing pairwise comparisons using the SSM server and the EMBOSS STRETCHER program. A plot of the percentage of C-α overlap at the reported RMSD versus RMSD (triangles) is also shown.

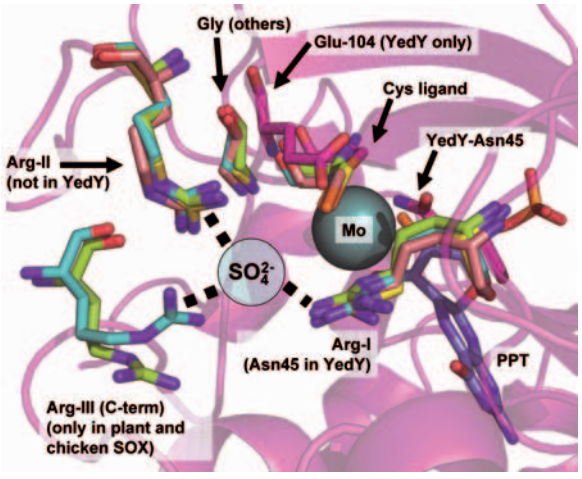

FIG. 6. Location of conserved residues within the active sites of the SUOX fold proteins. The five protein structures were structurally aligned using a combination of SSM server output and the ALIGN function of the PYMOL molecular graphics package. The plant (PDB accession number 1OGP) and chicken (accession number 1SOX) enzymes each contain the three Arg residues of the Arg triad, and these appear to stabilize the substrate (SO_3^{2-}) and/or product (SO_4^{2-}) in the active site. *Pichia* NIA (accession number 2BII) contains only one conserved active-site Arg, and YedY (accession number 1XDQ) contains none.

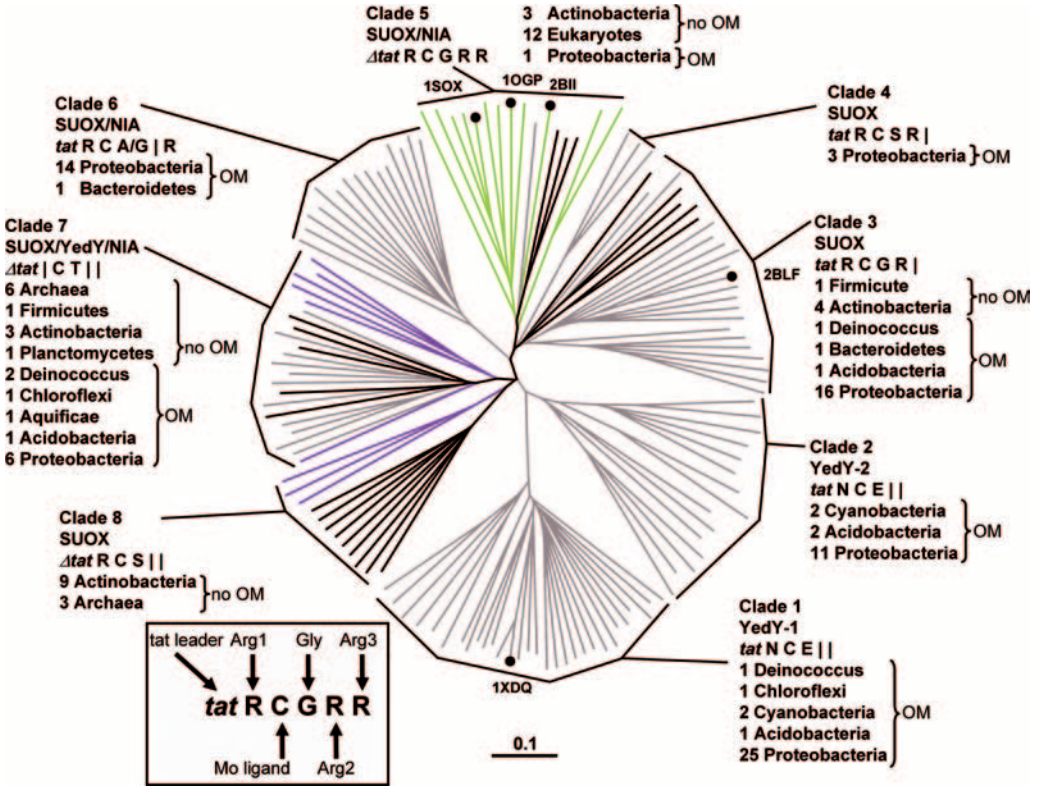

FIG. 7. Dendrogram of SUOX fold-containing protein sequences. As described in the text, sequences were mined from the UNIPROT database using the structurally characterized enzymes as bait. The resultant data set was aligned using the CLUSTALX program (87, 88) utilizing the Gonnet protein weight matrix (25), and obvious outliers were removed by inspection. These sequences were further filtered to eliminate sequences with greater than 60% identity to any other sequence in the set (using a script based on the EMBOSS STRETCHER program). This generated a final data set of 137 sequences that were analyzed to generate the dendrogram. These were aligned, and the figure was generated using the program TREEVIEW. To validate our dendrogram, we bootstrapped our alignment and generated a neighbor-joining tree (19, 76), which generated a distribution of clade membership identical to that presented in the figure. Default alignment parameters were used by the CLUSTALX program to generate the alignment. Sequences for which structures are available are marked with a "●". Putative functions were assigned by searching each sequence against the curated SWISSPROT database and are speculative. Branches appearing in gray are from diderm bacteria, those appearing in black are from monoderm bacteria, those appearing in blue are from archaea, and those appearing in green are from eukaryotes.

program) (70). This approach yielded a data set comprising 137 distinct sequences. Dendrograms grouping the SUOX-related proteins into discrete clades were generated using the TREEVIEW program (http://taxonomy.zoology.gla.ac.uk/rod/treeview.html; R. D. M. Page). In order to simplify this process, the clades were assigned on the basis of the simplified classification scheme outlined above (see Active-Site Sequence Conservation Suggests a Simplified Classification of SUOX Fold Enzymes), generating the unrooted tree presented in Fig. 7. Sequences corresponding to members of the individual clades were subjected to individual CLUSTALW alignments. Tentative functional assignments for each branch were based on BLASTP searches carried out against the annotated SWISSPROT protein sequence database hosted at the National Center for Biotechnology Information (at http://www.ncbi.nlm.nih.gov/) using the BLASTCL3 network client (available at http://www.ncbi.nlm.nih.gov/), which accesses the NCBI BLASTP server.

Inspection of Fig. 7 in combination with the simplified SUOX fold classification described above (see Active-Site Sequence Conservation Suggests a Simplified Classification of SUOX Fold Enzymes) enabled us to split the tree into eight distinct clades. The taxonomies of the organisms corresponding to each protein were extracted for each sequence from the UNIPROT flat-file entry. Our final data set included sequences from 77 *Proteobacteria* species, 19 *Actinobacteria* species, 11 eukaryote species, 9 *Archaea* species, 4 *Cyanobacteria* species, 4 *Deinococcus* species, 2 *Acidobacteria* species, 2 *Chloroflexi* species, 2 *Firmicutes* species, 1 *Aquificae* species, 1 *Bacteriodetes*, and 1 *Planctomycetes* species. Thus, the SUOX fold appears in organisms encompassing the entire span of evolution in the three domains of life.

One characteristic of the SUOX fold proteins that will be discussed in more detail below (see Correlation between the Presence of a *tat* Leader Sequence and a Diderm Bacterial Cell Envelope Morphology) is the correlation between clade assignment, the presence of a *tat* leader, and bacterial cell envelope morphology. Because a significant number of species possessing outer membranes are reported as lacking one on the basis of the Gram stain, we are reluctant to use the "gram positive"/"gram negative" terminology herein. We instead choose to use the terms monoderm (lacking an outer membrane) and diderm

TABLE 2. Conserved motifs within individual clades of proteins containing the SUOX fold

Clade[a]	No. of sequences[b]	Function[c]	Simplified classification[d]	Arginine triad (no. of arginine residues)[e]	Dimerization domain[f]	Heme binding domain[g]	FAD domain[h]	Conserved motifs[i]
1	30	YedY-1	*tat* N C E | |	No	No	No	No	NFYEF, RCVExW, PYxExL, **GAPxR**, PWKYGFK, YGF xANVNP, NGY
2	15	YedY-2	*tat* N C E | |	No	No	No	No	CVEGW, HPQT, KLGYK, **GxPxR**
3	24	SUOX	*tat* R C G R |	Yes (2)	Yes	Yes (subunit)	No	CxGNxR[9], **GxPxR**
4	3	SUOX	*tat* R C S R |	Yes (2)	Yes	Yes (subunit)	No	CSGNGR, **GxPxR**[j]
5	16	SUOX, NIA	Δ*tat* R C G R R	Yes (3)	Yes	Yes/no	Yes/no	PxNxEP, RNH, CAGNRR, H**GxPxR**, GARxVKW (I/L), RV(E/D)(V/L)S, WNxRxG
6	15	SUOX, NIA	*tat* R C (A/G) | R	Yes (2)	Yes	Yes (subunit)	No	RxHxG, HGxV, FxECxxN, WTGV, EgxDx(S/A), QNGE, EQ**GYPxR**, GLA WSGxG, SRxxDxTGY
7	22	SUOX, YedY, NIA	Δ*tat* | C T | |	No	No	No	No	FHCVTxWS, **GxPxR**, GYWE
8	12	SUOX	Δ*tat* R C S | |	Yes (1)	Yes	No	No	DFYR(V/I)D, CVSN, H**GY PVR**, GxAWA, PxGA (S/T)G

[a] Branches identified in Fig. 7.
[b] Number of sequences in each clade of Fig. 7.
[c] Function identified by BLAST searches of sequences against the curated SWISSPROT database.
[d] Simplified sequence classifications as described in the text. This is the basis for assignment to the individual clades shown in Fig. 7.
[e] The arginine triad stabilizes the product/substrate at the active site of SUOX (47). The number of Arg residues is given in parentheses.
[f] The dimerization domain, if present, is located at the C terminus of the SUOX domain.
[g] The heme binding domain can either comprise a separate subunit (as in the SUOX from *S. novella*, clade 3), can comprise a separate domain at the N terminus of the SUOX fold (as in plant SUOX, clade 5), or can be located C terminal to the dimerization domain (as in plant NIA, clade 5).
[h] The FAD domain is located at the C terminus of the plant-type nitrate reductase.
[i] Underlined motifs imply conservation across a few clades. Motifs in boldface type imply conservation across all clades.
[j] Motifs for clade 4 are underreported. As clade 4 contains only three members, the sequences are highly conserved across their entire lengths. As such, only common motifs found in other clades are reported therein.

(possessing an outer membrane) (29). Species with outer membranes were identified on the basis of morphological studies or the presence of readily identifiable outer membrane proteins encoded by their genomes. In the context of this work, the following domains/phyla were deemed to be monoderms: *Archaea*, *Firmicutes*, *Actinobacteria*, and *Planctomycetes*. The following bacterial phyla were deemed to be diderms: *Proteobacteria*, *Deinococcus*, *Chloroflexi*, *Cyanobacteria*, *Bacteriodetes*, *Aquificae*, and *Acidobacteria*.

Diderm bacterial species are gray in Fig. 7, while monoderms are black. Eukaryotic species are green, and the *Archaea* are blue. It is clear that the archaeal sequences are located in clades 7 and 8, whereas the eukaryotic species are located in clade 5. Each clade is distinguished by a group of conserved motifs listed in Table 2 and discussed below.

Clade 1: Bona Fide YedY Sequences (YedY-1)

We have arbitrarily set clade 1 to be the branch containing YedY from *E. coli* and its close homologs. Sequences in this clade are derived from diderm species, with all the sequences possessing a *tat* leader (1, 78, 79, 91). Since in the case of *E. coli*, YedY is a soluble periplasmic enzyme, there is presumably an evolutionary advantage in having an outer membrane to prevent its loss to the external environment. The correlation between the presence of soluble *tat*-exported proteins and bacterial cell wall morphology has been noted elsewhere (75) and will be discussed in more detail below (see Correlation between the Presence of a *tat* Leader Sequence and a Diderm Bacterial Cell Envelope Morphology). Clade 1 is composed mainly of sequences from the *Proteobacteria*, with 24 entries, along with two for the *Cyanobacteria*, one for the *Acidobacteria*, one for *Deinococcus*, and one for *Chloroflexi*.

The sequence motifs conserved in the YedY-1 clade are indicated in Table 2 and will be discussed in detail below (see "Conserved Motifs in YedY").

Clade 2: YedY-2 Sequences That Are Closely Related to YedY-1

Clade 2 diverges from clade 1 but is very similar to it. It is distinguished by the presence of two unique conserved motifs, HxQT and KLGYK (Table 2). This branch is comprised mostly of *Proteobacteria*, with 11 entries in total. The remaining four homologs are found in two *Acidobacteria* sequences and two *Cyanobacteria* sequences. As is the case for clade 1, these organisms are all diderms, and each protein is predicted to be a *tat*-exported soluble periplasmic protein.

Clade 3: *tat*-Exported Periplasmic SUOX

Clade 3, along with clade 7, contains sequences from both monoderm and diderm bacterial species. The majority (14/24) of the sequences in clade 3 possess a *tat* leader peptide, indicating that these are secreted into the periplasm or extracellular space. Members of clade 3 are identified as being bacte-

rial SUOX enzymes in BLASTP searches against the SWISSPROT database and include the structurally characterized SorAB SUOX of *S. novella* (41). Each member of the clade possesses the first two Arg residues of the Arg triad. Members of this clade containing an outer membrane are 16 *Proteobacteria*, one *Acidobacteria*, one *Deinococcus*, and one *Bacteriodetes* species. Monoderm members comprise four *Actinobacteria* sequences and one *Firmicutes* sequence.

Clade 4: a Second Clade of *tat*-Exported Periplasmic SUOX

Clade 4 contains the fewest members of those represented in Fig. 7 and has a distinct simplified classification (Table 2). Its three members all belong to the *Proteobacteria*. Like clade 3, these sequences possess the first two Arg residues of the Arg triad.

Clade 5: Eukaryotic SUOX and NIA Enzymes

Clade 5 includes the archetypal eukaryotic SUOX and NIA enzymes, including plant SUOX (PDB accession number 1OGP) (82), chicken SUOX (accession number 1SOX) (47), and *P. angusta* NIA (accession number 2BII) (22). All members of this clade lack a *tat* leader but contain the catalytic Arg triad. This clade is comprised of 12 eukaryotic proteins, six of which belong to the *Metazoa*, three of which belong to the *Fungi*, two of which belong to the *Viridiplantae*, and one of which belongs to the *Mycetozoa*. The remaining species belong to the *Actinobacteria*, with three entries, and the *Proteobacteria*, with one entry. As will be discussed below, it is notable that this clade contains sequences in which there is a variability of domain order. For example, in chicken SUOX, the heme b_5 domain precedes the SUOX fold, but in the plant-type nitrate reductases (including that of *P. angusta*), it follows the dimerization domain. It is absent in the plant-type SUOX. These observations suggest that "domain shuffling" may have occurred recently on the evolutionary timescale.

Clade 6: Bacterial Periplasmic SUOX and NIA Enzymes

Clade 6 encompasses enzymes tentatively identified in BLASTP searches against the curated SWISSPROT database as bacterial SUOXs and NIAs. Fourteen *Proteobacteria* and one *Bacteriodetes* species are represented. All the homologs possess a *tat* leader and are predicted to be exported into the periplasm. In this group, the first and third Arg residues of the Arg triad are conserved.

Clade 7: Cytoplasmic Bacterial SUOX/YedY/NIA

The bacterial species observed in clade 7 are the most diverse of those represented in Fig. 7. The sequences lack a *tat* leader, and they also completely lack the Arg residues of the Arg triad. On this basis, although tentatively identified functionally as being SUOX/YedY/NIA enzymes, it is unlikely that they interact with simple inorganic anionic substrates. Species lacking an outer membrane are represented by six *Archaea*, three *Actinobacteria*, one *Firmicutes*, and one *Planctomycetes* species. Diderm bacterial species are represented by six *Proteobacteria*, two *Deinococcus*, one *Chloroflexi*, one *Aquificae*, and one *Acidobacteria* species. The presence of archaeal sequences indicates that this clade arose very early on in evolution.

Clade 8: Cytoplasmic SUOX Enzymes from Monoderm Bacteria

Uniquely, clade 8 exclusively contains sequences from monoderm bacteria and *Archaea*. The first Arg of the Arg triad is retained, and the active-site Gly is replaced by a Ser residue. Nine sequences are from the *Actinobacteria*, and three are from the *Archaea*. The lack of a *tat* leader indicates that the mature proteins are cytoplasmically localized.

PRESENCE AND FUNCTION OF THE DOMAINS OBSERVED IN STRUCTURALLY CHARACTERIZED PROTEINS CONTAINING THE SUOX FOLD

Three structurally identifiable domains are observed in the eukaryotic SUOX/NIA clade (clade 5) (Fig. 7). These are the critical Mo-PPT binding SUOX domain, the heme b_5 domain, and the dimerization domain. Within the subgroup of proteins of known structure, these are best represented by the SUOX enzyme from *G. gallus* (47). In *G. gallus* SUOX, the domain order is heme b_5, Mo-PPT, and then dimerization. In the plant-type NIA, the domain order is Mo-PPT, dimerization, heme b_5, and then an FAD-containing NADH-oxidizing domain (8). The structure of the NADH-oxidizing flavoprotein domain from maize has been solved to a resolution of 2.5 Å (PDB accession number 1CNF) (54). As can be seen in Fig. 1 and 2, the SUOX fold is common to all members and defines the family. The variability of the presence and locations of the other domains within the SUOX fold enzymes can provide important clues as to the evolutionary relationships among the sequences represented in Fig. 7.

The dimerization domain can act as the interface between SUOX fold monomers, and many enzymes containing the SUOX fold have been shown to be dimers (20). However, the role of the dimerization domain in enzyme function remains enigmatic. While the monomeric forms of SUOX enzymes are catalytically active, the dimerized form can be threefold more active (12). Mutations that cause an SUOX deficiency in humans have been identified in the dimerization domain as well as in the SUOX fold, highlighting the importance of the latter (46).

As will be discussed below (see Evolutionary Pathway of the SUOX Fold Proteins), the appearance of a heme-containing domain or subunit appears to be a critical evolutionary step following the emergence of the dimerization domain. In the case of *S. novella* SUOX (SorAB), a separate heme *c*-containing subunit (SorB) is observed in the protein structure (PDB accession number 2BLF) in a conformation that provides a center-to-center distance (Fe to Mo) of 16.6 Å (41). Crucially, the edge-to-edge distances between these cofactors are only 7.0 Å from one of the heme propionates to the PPT and 8.5 Å from the same propionate to the Mo atom of the Mo-PPT. Kappler et al. (41, 42) identified a plausible pathway of bond-orbital and H bonds that could serve as a conduit between the two redox-active cofactors.

In the case of chicken SUOX, the situation is quite different. The heme-containing b_5 domain is located at the N terminus of

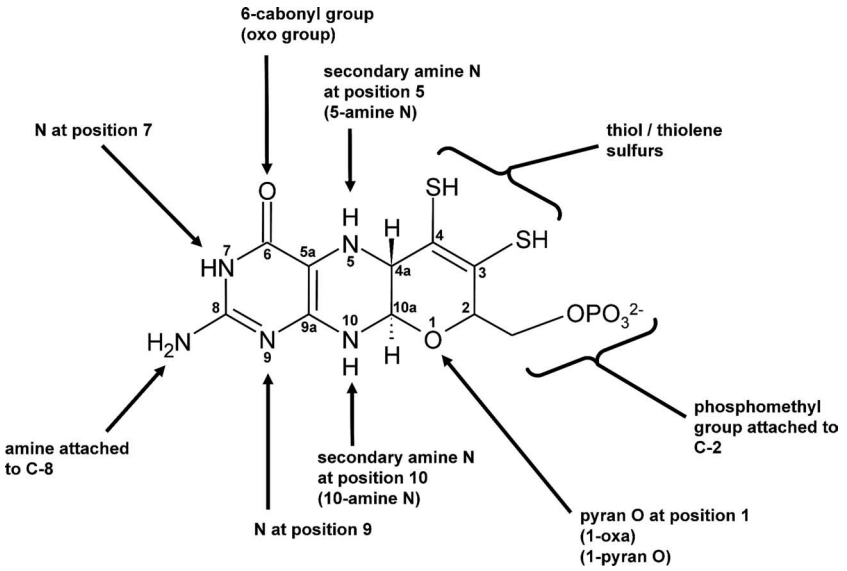

FIG. 8. Structure of the PPT moiety of the Mo-PPT cofactor. The correct naming conventions for individual atoms of the PPT are detailed. This nomenclature is used for interactions between the protein and the PPT discussed in the text.

the protein, but the edge-to-edge distance between the two cofactors is 21.3 Å, and the center-to-center distance is 25.8 Å (47). These distances are beyond the 14-Å maximum that has been proposed to be the maximum for kinetically competent intercenter electron transfer (61–64, 68). It has been proposed that there is a hinge region comprising approximately 10 amino acids between the heme b_5 domain and the Mo-PPT/dimerization SUOX fold that permits domain rearrangement, allowing the heme to approach the Mo-PPT to permit catalytically competent electron transfer rates (67). This hypothesis is based on the heme domain interacting electrostatically with the Arg triad in the absence of substrate or product, thus facilitating interdomain electron transfer.

In the case of NIA enzymes such as that from *P. angusta*, the presence of a similar hinge region between the dimerization domain and the heme b_5 domain that is localized C terminal to the dimerization domain has been proposed (8). Indirect evidence for the role of this hinge in activating plant NIA has been obtained from studies using spinach (*Spinacea oleacea*) NIA as a model system. This enzyme shares the same domain order within its primary sequence as that observed in the *P. angusta* enzyme, with the heme b_5 domain following the SUOX fold. Spinach NIA is inactivated in the dark via a process in which a Ser residue (Ser543) in the hinge region connecting the Mo-PPT dimerization domain with the heme b_5 domain is phosphorylated (16, 60), followed by binding of the NIA inhibitor protein (24, 55, 60). Binding of the NIA inhibitor protein presumably constrains the hinge function and results in a Mo-PPT-to-heme distance well beyond the 14-Å limit for catalytically competent intercenter electron transfer. Thus, in both animal SUOX and plant NIA enzymes, available data suggest a role for hinge regions in facilitating and regulating interdomain electron transfer. Interestingly, heme *c* of *S. novella* SUOX is located in close juxtaposition not only to the Mo-PPT cofactor but also to the substrate binding funnel, suggesting that perhaps its location is functionally equivalent to that of the heme b_5 domain of chicken SUOX (67).

PROTEIN-PPT INTERACTIONS

Figure 8 shows the structure and selected nomenclature of the organic moiety of the Mo-PPT cofactor that is bound by the SUOX fold. It is the product of a complex biosynthetic pathway that has been extensively reviewed (57, 58). It incorporates a large number of noncarbon atoms able to interact with a range of side chain functionalities of the protein scaffold. As discussed below, many of the conserved motifs identified in Table 2 contribute to Mo-PPT coordination in a clade-specific manner.

It is notable that the E_m values for the Mo(IV/V) and Mo(V/VI) couples vary considerably between the enzymes sharing the SUOX fold. In chicken SUOX, the E_m values have been reported to be −86 mV and 131 mV for the Mo(IV/V) and Mo(V/VI) couples, respectively (86), and in *S. novella*, they are reported to be −46 mV and 160 mV (42). Thus, a partial consensus suggests that the true SUOXs have widely separated E_m values that render the Mo(V) electron paramagnetic resonance spectrum readily observable. In the case of the NIA enzymes, the potentials are closer together: the reported values for the spinach enzyme (*Spinacea oleacea*) are −6 mV and 2 mV (45), and in green algae (*Chlorella vulgaris*), the reported values are −54 mV and −34 mV (84). Thus, in the NIA enzymes, the Mo(V) is less easily observed by electron paramagnetic resonance than are true SUOX enzymes. Finally, YedY from *E. coli* exhibits only the Mo(IV/V) transition, and this occurs with an E_m of 132 mV (7). Thus, in YedY, the Mo(V) state is the most accessible of all the SUOX fold enzymes studied to date. Because of the observed complexities of the Mo coordination environment in enzymes having the SUOX fold (20, 34, 35, 37, 46), the factors determining the

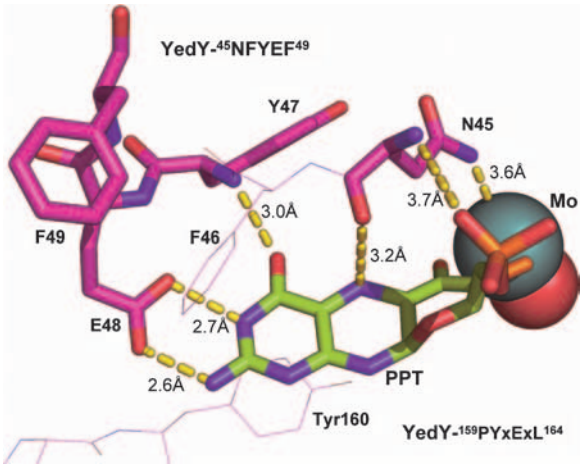

FIG. 9. Active-site motifs of YedY. Conserved motifs uncovered by the CLUSTALX alignment were analyzed in the structure of YedY. The conserved motif NFYEF from clade 1 makes six potential interactions with Mo-PPT. The conserved motif PYxExL contributes an aromatic ring (Tyr160 in YedY) to an aromatic base-stacking formation consisting of F45, Y160, and the outer ring of the pterin group. The cysteine-containing motif is not visualized in this figure. All distance values are within the accepted values for hydrogen bonding.

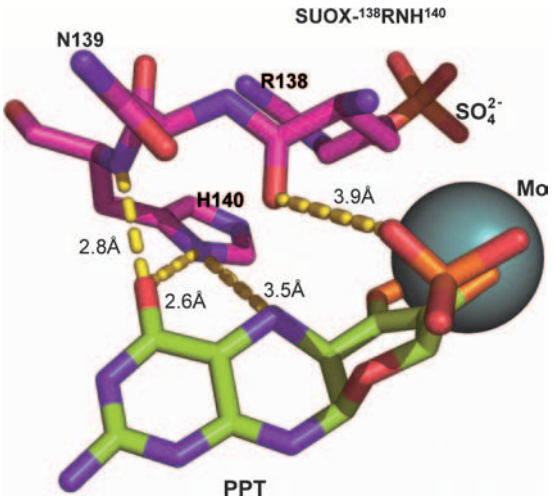

FIG. 10. RNH motif of chicken SUOX. Conserved motifs uncovered by the CLUSTALX alignment were analyzed in the structure of chicken SUOX. The RNH motif found in the active site of clade 5 enzymes contributes to the coordination of the PPT (four potential interactions), including R138, a crucial residue in the Arg triad. The product of sulfite oxidation, sulfate, is in close proximity to R138.

Mo(IV/V) and Mo(V/VI) midpoint potentials are poorly understood. It is possible that the coordination of the organic component of the Mo-PPT cofactor may play a crucial role in defining the Mo(IV/V) and Mo(V/VI) midpoint potentials.

Conserved Motifs in YedY

The conserved motifs of *E. coli* YedY (of the YedY-1 clade) are concentrated mostly around the PPT cofactor (53), and these motifs participate primarily in its coordination. Of primary importance is ^{45}NFYEF^{49} (Fig. 9). Asn45 interacts with the phosphate group of the PPT via its backbone amide nitrogen (3.7 Å). Its side chain nitrogen also interacts with the Mo atom (3.6 Å). The backbone amide nitrogen of Tyr47 hydrogen bonds to the 6-keto oxygen of the pterin. Glu48 appears to stabilize the 7-amine nitrogen as well as the amine attached to C-8 of the pterin using both of its side chain carboxylate oxygens. This Glu residue is not conserved in any of the other clades presented in this study. An additional conserved motif, ^{159}PYxExL^{164}, is also unique to the YedY-1 clade, and some detail of this is also shown in Fig. 9. While most of its residues do not interact with Mo-PPT, Tyr160 appears to "stack" with the PPT ring system in a manner similar to that observed between a Tyr residue (Tyr1005) and one of the pterin rings of the Mo-bisPGD cofactor of FdnGHI (40). A similar phenomenon occurs with the Tyr47 residue of the YedY-1 ^{45}NFYEF^{49} motif, but in this case, there is a greater interplane angle between the Phe ring and the two coplanar rings of the PPT. The side chains of both Tyr47 and Tyr160 are located within ~4 Å of the PPT ring system. Interestingly, the absolutely conserved Lys207 serves an essential role in the coordination of the PPT, as it interacts with the C-8 amine and 9-imino nitrogen of the pterin ring as well as with two oxygens of the phosphate group (Fig. 4). This Lys, while absolutely conserved within the structurally characterized SUOX fold proteins, is not surrounded by other conserved residues within an identifiable motif.

Three additional motifs also appear in the general area of the Mo-PPT active site of YedY. These motifs are ^{101}RCVEx W^{106}, ^{198}PWKYGFK^{204}, and ^{231}YGFxANVNP^{239}. Notwithstanding the role of the conserved Cys residue in Mo coordination and catalysis, Tyr231 (7.2 Å), Tyr47 (7.6 Å), and the previously mentioned Asn45 (3.6 Å) appear to be close enough to the Mo and its adjacent substrate binding site to have an influence on the catalysis of the as-yet-unknown interconversion catalyzed by members of the YedY-1 clade.

Another YedY-1 motif identified in Table 2, ^{271}NGY^{273}, is found on the surface of YedY. While the crystal structure of YedY was found as a pentamer, the ^{271}NGY^{273} domain does not occur between monomer surfaces. The ^{190}GAPIR^{194} motif, which is found conserved as GxPxR across all of the clades, is discussed below (see "Structural Motifs in Clades of Unknown Structure and Function"). Overall, we conclude that at least two of the conserved motifs of the YedY-1 clade are involved in PPT coordination.

Conserved Motifs in *P. angusta* NIA and Plant and Chicken SUOX (Clade 5)

The known structures of proteins represented in clade 5 contain motifs that serve a function similar to those found in clade 1 but which have distinct sequences. In the following text, residues are numbered according to chicken SUOX numbering, although the motifs are also conserved in plant SUOX and *P. angusta* NIA. The function of the YedY-1 ^{45}NFYEF^{49} motif is assumed by the ^{138}RNH^{140} motif in the clade 5 enzymes. While YedY-1 ^{45}NFYEF^{49} appears to interact with Mo-PPT at up to six positions (Fig. 9), the ^{138}RNH^{140} motif appears to make only two interactions (Fig. 10). The first is with the PPT 6-keto oxygen group via a potential hydrogen bond between the 6-keto oxygen and the backbone amide nitrogen of the

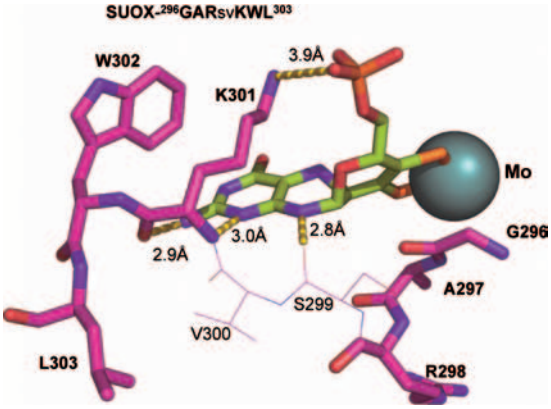

FIG. 11. GARsvKWL motif of chicken SUOX. The conserved motifs uncovered by the CLUSTALX alignment were analyzed in the structure of chicken SUOX. The GARsvKWL motif is shown here, making multiple interactions with the PPT group as well as providing a conserved environment for the molybdenum cofactor (Mo). The Lys residue in position 301 is found to be conserved in all of the known structures of SUOX fold proteins, always contributing two interactions to the outer ring of the pterin and one interaction to the phosphate group.

histidine (2.8 Å). The 6-keto oxygen is also only 2.6 Å from one of the imidazole nitrogens of the His140 side chain and could hydrogen bond with it in its protonated form. The backbone carbonyl oxygen of Arg138 is located 3.9 Å from one of the phosphate oxygens. Note that Arg138 is equivalent to Arg1 of the Arg triad (Fig. 6).

A second motif, ^{296}GARxxKW(I/L)^{303}, also interacts with the PPT (Fig. 11) but on the opposite side of the ring system compared to the interactions observed with the ^{138}RNH^{140} motif. The ^{296}GARxxKW(I/L)^{303} motif includes the absolutely conserved Lys residue equivalent to YedY-K207 (Fig. 4). The amide bond nitrogen of this Lys301 interacts with the N-9 of the PPT ring system (Fig. 8). Its side chain amine interacts with the PPT phosphate group. The amide bond carbonyl oxygen interacts with the amine attached to the C-8 position in the PPT ring system. These three interactions are absolutely conserved among the structurally characterized SUOX fold enzymes. Of the other residues within the ^{296}GARxxKW(I/L)^{303} motif, it is the backbone carbonyl oxygen of the first "x" residue (Ser299) that appears to be of greatest importance in coordinating the PPT. Its carbonyl oxygen appears to form a hydrogen bond with the N-10 nitrogen.

In the clade 5 proteins, the ^{185}CAGNRR^{190} motif is equivalent to the YedY-1 ^{101}RCVExW^{106} motif (Fig. 3 and Table 2). In both cases, the Cys residue is the protein-Mo ligand, and in the clade 5 ^{185}CAGNRR^{190} motif, the second Arg residue (Arg190) is Arg2 of the Arg triad (Fig. 3 and 6). The clade 5 ^{114}PxNxEP^{119} motif functions to coordinate a portion of the ^{138}RNH^{140} motif (not shown), with a hydrogen bond between the side chain oxygen of the ^{114}PxNxEP^{119} Asn and one of the side chain nitrogens of the ^{138}RNH^{140} Arg. When looking at a surface representation of chicken SUOX, the ^{138}RNH^{140} motif is found at the active-site funnel (not shown). Additionally, the ^{114}PxNxEP^{119} motif is located adjacent to it. The final conserved motif of this family, ^{381}RV(E/D)(V/L)S^{385}, is found in the dimerization domain.

Conserved Motifs in *S. novella* SUOX

Clade 3, containing *S. novella* SUOX, shows a marked lack of conserved motifs compared to the other clades. In this case, only two motifs can be found (Table 2), and they resemble motifs of the previously described clades. Due to the lack of motifs, analysis of individual conserved residues was performed. In many cases, the conserved residues were related to those found in the other clades as well, providing insight into exactly which residues are of importance in each motif. As is the case with the other clades containing members of known structure, the active-site Cys is found in the same position, as well as an active-site Tyr and the Lys bearing multiple interactions with the PPT. While the *P. angusta*, plant, and chicken enzymes all have a conserved motif in their dimerization domains, the *S. novella* SUOX has only three absolutely conserved residues in the dimerization domain. These residues (Trp307, Gly346, and Gln349), while not in an identifiable motif, are localized to the protein surface. Additional semiconserved Trp and Leu residues are found in this region. However, these residues are not found on the surface of the dimerization domain. There are two additional conserved Gly residues, which are not found near the surface, but these may be to add flexibility to the dimerization domain, as they are both found near the "joint" of this domain.

Structural Motifs in Clades of Unknown Structure and Function

Across all known structures, there exists a surprisingly low level of conservation of residues surrounding the active site (Fig. 4). The conserved motifs of the remaining branches can be seen in Table 2. Unfortunately, no crystal structures are available for enzymes found in these branches, and as such, an analysis of the potential functions of these motifs cannot be done. This problem is compounded by a relative lack of experimental data on these enzymes.

Aside from the integral Cys residue, all of the known structures contain a Lys residue that coordinates the PPT as well as an Arg residue that interacts with the phosphate group of the PPT (Fig. 4). This Arg is part of a motif (GxPxR) that is conserved across all of the clades of our analysis. However, most branches differ in regard to the residues preceding and internal to this motif. Table 2 shows that in clades 5 and 8, GxPxR is preceded by a His. In the structures of clade 5 enzymes, this His hydrogen bonds to the same phosphate oxygen that is being coordinated by the Arg of the GxPxR motif. In YedY, as well as most of the clade 1 enzymes, the GxPxR motif is usually preceded by an Asn. Much like the His residues, this Asn interacts with the phosphate group. The backbone carbonyl group of this residue also appears to interact with the side chain oxygen of the Arg. The Gly and Pro that are internal in this motif most likely give the motif a characteristic "U shape." The Pro in the center of the motif causes a turn in the backbone, while the Gly gives freedom to the preceding residue, allowing it to adopt a proper orientation. Thus, it would appear that the GxPxR motif offers extensive coordination of the phosphate group.

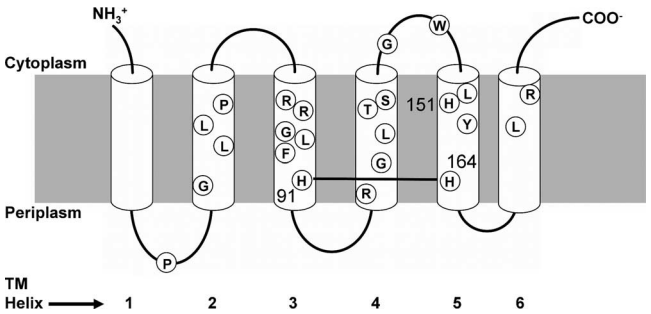

FIG. 12. Predicted TM topology and distribution of conserved residues in the YedZ family. A sequence data set of 107 YedZ sequences was subjected to hydropathy analyses as described in the text. The positions of the absolutely conserved residues within the bacterial YedZ family are shown. The numbering of the TM helices used in the text is indicated beneath the topology model.

DIFFERENTIAL REDOX PARTNERS

The redox partners of the clade 5 enzymes appear to be well established. As described above, the heme b_5 domain of animal SUOX appears to swivel toward the Mo-PPT active site to accept electrons from the Mo-PPT cofactor. Plant SUOX is localized to the peroxisomes and appears to donate electrons directly to oxygen to form peroxide (32). Plant NIA also appears to undergo domain movement to facilitate electron transfer to its heme. In the case of the structurally characterized clade 3 SorAB enzyme from *S. novella*, an accessory subunit that coordinates a heme *c* appears to be the primary electron acceptor from the Mo-PPT active site, with no domain or subunit movement being required for enzyme turnover. Little is known about the redox partners of members of the remaining clades of SUOX fold enzymes.

As stated above, the second gene in the *E. coli yedYZ* operon, *yedZ*, encodes a TM heme-containing protein (7). In order to gain further insights into YedZ, *E. coli* YedZ was used as bait in a BLASTP search of the UNIPROT database, and we obtained a sequence data set comprising 107 sequences after filtering pairwise with a 95% identity cutoff filter. These sequences were aligned using CLUSTALX (87, 88), and the resultant alignment was analyzed using the EMBOSS TMAP program (70) to generate the TM topology shown in Fig. 12, which is essentially in agreement with that proposed by Drew et al. (17). Inspection of Fig. 12 reveals the placement of sequential Arg residues (Arg77 and Arg78) toward the cytoplasmic end of TM helix 3. In order to verify the predicted inclusion of these residues within the hydrophobic core of YedZ, we subjected the entire sequence set to analysis using the TMHMM server (version 2; https://www.cbs.dtu.dk/services/TMHMM/) (50, 85). The TM topology predicted for the YedZ sequence set by the TMHMM server is essentially identical to that predicted by the EMBOSS TMAP program.

Twenty-four residues are highly conserved within the YedZ family, including the two Arg residues mentioned above. Most significant are three His residues that are likely candidates for heme-iron coordination. His91 (*E. coli* numbering) and His164 are located toward the periplasmic ends of TM helices 3 and 5, and these are the most likely to be involved in the coordination of heme *b* observed by Brokx et al. (7). Hydrophobic *b*-type cytochromes typically have conserved Arg residues whose side chains interact with one or more of the heme propionates. Examples of this include mitochondrial and bacterial complex II (93), *E. coli* nitrate reductase A (4, 5), and the cytochrome bc_1 complex (94). In the case of YedZ, a candidate Arg for interacting with the heme propionates is located toward the periplasmic end of TM helix 4 (Arg115). The identification of residues involved in quinol binding and redox reactions is more problematic, with a cluster of conserved residues being located at the cytoplasmic ends of TM helices 4 and 5, including His151 and the nearby Tyr155.

Of the 107 YedZ sequences obtained in our analyses, 98 of the sequences are proteobacterial in origin, 5 are from the *Chloroflexi*, 2 are from the *Acidobacteria*, and two are from the *Deinococci*. All these species are diderms. When sequences are mined using *E. coli* YedY as bait, a data set of 114 sequences is obtained following the use of a 95% cutoff filter. Of these, 103 are proteobacterial, 5 are from the *Chloroflexi*, 4 are from the *Cyanobacteria*, and two are from the *Deinococci*, and all the phyla represented are diderms. Thus, with the exception of the representation of cyanobacterial species in the YedY sequence set, there is a strong correlation between the phylogenic representations of YedY- and YedZ-type subunits, with both sets being largely proteobacterial in origin. In the case of YedY, each sequence bears a *tat* leader at its N terminus, and the status of this subunit as a soluble periplasmic enzyme renders the presence of an outer membrane necessary to prevent its loss to the bulk milieu.

CORRELATION BETWEEN THE PRESENCE OF A *tat* LEADER SEQUENCE AND A DIDERM BACTERIAL CELL ENVELOPE MORPHOLOGY

The correlation between the observation of YedY sequences and a diderm cell envelope morphology of the species in which they are found prompted us to investigate the correlation between the localization of bacterial SUOX-type enzymes and the presence of an outer membrane. We examined this issue in each of the clades shown in Fig. 7, with clades possessing a *tat* leader being deemed to contain periplasmically localized proteins. Table 3 shows the results of our analyses of the species encoding the 137 sequences represented by the eight clades shown in Fig. 7. There is a strong correlation between the presence of a *tat* leader and the sequence coming from a diderm species. This is clearly demonstrated by clades 1, 2, 4, and 6. In each case, all of the sequences come from species possessing an outer membrane. In clade 3, 19 out of 24 sequences come from diderm species, with 14 out the 24 sequences having *tat* leaders. The monoderm species of clade 3 are four actinobacterial species (*Arthrobacter aurescens*, *Saccharopolyspora erythraea*, *Streptomyces ambofaciens*, and *Rubrobacter xylanophilus*) and one *Firmicutes* species (*Bacillus* sp. strain NRRL B-14911).

The clades dominated by sequences lacking a *tat* leader have a more balanced distribution between diderm and monoderm species. Species represented in clade 7 are equally distributed between the cell envelope morphologies, and those represented in clade 8 are exclusively from monoderm species (*Actinobacteria*). The proteins represented in clades 7 and 8 are predicted to be soluble cytoplasmic enzymes, and thus, no

TABLE 3. Correlation between periplasmic localization and presence of an outer membrane

Clade[a]	Function[b]	Total no. of species[c]	No. of species containing an OM[d]	*tat* leader[e]	Major group (no. of species)
1	YedY-1	30	30 (0)	Yes	Proteobacteria (25)
2	YedY-2	15	15 (0)	Yes	Proteobacteria (11)
3	SUOX	24	19 (5)	Yes	Proteobacteria (16)
4	SUOX	3	3 (0)	Yes	Proteobacteria (3)
5	SUOX/NIA	16	1 (15)[f]	No	Eukaryotes (12)
6	SUOX/NIA	15	15 (0)	Yes	Proteobacteria (15)
7	SUOX/YedY/NIA	22	11 (11)	No	Archaea (6), proteobacteria (5)
8	SUOX	12	0 (12)	No	Actinobacteria (9)
Total	SUOX fold	137	93		

[a] Clades are indicated in Table 2 and Fig. 7.
[b] Function identified by BLAST searches of sequences against the curated SWISSPROT database, except in cases where function has been assigned by experiment.
[c] Total number of species represented by each clade shown in Fig. 7.
[d] Species from bacterial phyla possessing an outer membrane (OM). Numbers in parentheses indicate the number of additional entries in each clade that are comprised of archaea, monoderm bacteria, and eukaryotes.
[e] The presence of a *tat* leader was determined by inspection of the sequence alignments of sequences assigned to each clade.
[f] Inclusion of clade 5 is done for the sake of completeness; however, four prokaryotic species are represented in it.

correlation between their occurrence and the cell envelope morphology of the species in which they are found is expected. The presence of an outer membrane in the species represented in clades 1, 2, 4, and 6 presumably serves to prevent the loss of the *tat*-targeted SUOX-type proteins to the bulk milieu. Similar correlations have been observed in analyses of other closely related families of proteins that can be either cytoplasmically located or targeted for export by the *tat* system (75).

CORRELATION BETWEEN THE PRESENCE OF SUOX FOLD PROTEINS AND YedZ

Our analyses of SUOX fold protein and YedZ sequence data prompted us to investigate the correlation between their occurrences within the genomes analyzed herein. This issue is important because YedZ may interact with a variety of non-SUOX-fold periplasmic proteins. We addressed this issue by comparing the unfiltered list of SUOX fold proteins (559 entries) with an unfiltered list of YedZ homologs (171 entries). All species containing both a SUOX fold protein and a YedZ homolog were compiled, and the former species were sorted into branches based on the presence of the conserved motifs described above. The results show a correlation between the appearance of YedZ homologs and the species represented in clades 1, 2, 4, and 6 (Fig. 7).

Utilizing an approach using both the accession numbers of the proteins and an operon searching server (http://string.embl.de/newstring_cgi/show_input_page.pl#), we identified an additional membrane-bound cytochrome *b* family whose genes appear adjacent to those of a subset of the SUOX fold proteins. This family, referred to herein as FdnI-like, is represented by Q2N9I4_ERYLH (UNIPROT accession no. for *Erythrobacter litoralis*), which has a TM topology similar to that found in the four TM members of the diheme FdnI family (40, 75). The distributions of the two types of hydrophobic cytochromes *b* that may interact with SUOX fold enzymes are listed in Table 4.

DISTRIBUTION OF ALTERNATIVE CYTOCHROME REDOX PARTNERS

Compared to many other redox enzyme systems, enzymes of the SUOX class present relatively simple electron transfer relay architectures. These enzymes do not contain chains of [Fe-S] clusters, and their relays typically encompass the Mo-PPT, a heme, and, in the NIA enzymes, an extra NADH

TABLE 4. Distribution of differential redox partners

Clade	No. of species with cell wall morphology of[a]:		No. of species of domain:			*tat*[b]	No. of species with cytochrome family of type[c]:						
							b		*c*				
	Monoderm	Diderm	Bacteria	Archaea	Eukaryotes		YedZ	FdnI	1	2	3	4	5
1		30	30			Yes	25						
2		15	15			Yes	8	5					
3	5	19	24			Yes	10	5	11	4	2	1	
4		3	3			Yes	2	1				3	
5	3	1	4		12	No							
6		15	15			Yes	5	3					13
7	11	11	16	6		No							
8	12		9	3		No							

[a] Numbers relate to the species represented in Fig. 7.
[b] *tat*, targeted to the *tat* translocon in the majority of cases.
[c] See the text for more information on the cytochrome families.

binding FAD domain. As mentioned above, a hinge region in animal SUOX enzymes appears to allow the heme b_5 domain to approach the SUOX fold domain and accept electrons from it. *S. novella* SUOX, rather than having a heme-containing domain, forms a heterodimer with a cytochrome *c* subunit (SorB). A preliminary BLAST search of the sequence of this SorB subunit returned hits identified as being other cytochrome *c* subunits. Akin to the YedZ-like protein analysis, we cross-referenced the accession numbers, taxonomies, and operon positions of the cytochrome *c* proteins against our phylogeny of SUOX fold proteins. In many cases, a cytochrome *c* gene could be found in an operon with a SUOX protein. We thus determined that only 11 members of our phylogeny contained a SorB-type cytochrome *c* subunit, and all of these were in clade 3. As seen in Table 2, not only members of clade 3 but also the members of clades 4 and 6 contain separate cytochrome *c* subunits that are associated with a SUOX fold protein. By searching by accession number and gene position, we identified an additional four unique classes of cytochromes *c*. These are exemplified by proteins with UNIPROT accession no. A5V4U9_9SPHN (*Sphingomonas wittichii* RW1), A4AQV2_FLAO (*Flavobacteriales bacterium*), A3SHJ5_9RHOB (*Roseovarius nubinhibens*), and Q9LAH4_THINO (*S. novella*). The distribution of these families across our phylogeny is indicated in Table 4. Thus, there are five families of cytochromes *c* that can be predicted to interact directly or indirectly with their respective SUOX fold proteins. It is also notable that these cytochromes *c* appear in operons with SUOX fold proteins bearing a *tat* leader, suggesting that they form part of electron transfer systems coupling the periplasmic SUOX fold proteins to the membrane-intrinsic quinone pool. However, with the exception of SorB (family 1), there is no confirmation that these proteins form stable heterodimeric complexes with their putative SUOX fold partners.

Four of the cytochrome *c* families appear in members of clade 3, which agrees with the observation that clade 3 contains a greater taxonomic diversity than clades 4 and 6 do. Both clades 4 and 6 claim a single cytochrome *c* family. The distribution of these cytochrome *c* proteins is shown in Table 4. Cytochromes *c* appear in three clades totaling 41 members. However, six of the members (five from clade 3 and one from clade 6) are missing this subunit.

PRESENCE OF MULTIPLE REDOX SYSTEMS

Above (see Correlation between the Presence of SUOX Fold Proteins and YedZ and Distribution of Alternative Cytochrome Redox Partners), we detailed the presence of both soluble and membrane-bound redox partners of the SUOX fold proteins. Interestingly, and as detailed in Table 4, these two systems overlap in several taxonomies. Inspection of our unfiltered list of SUOX fold proteins (559 entries) reveals that many species contain multiple (at least three) SUOX fold proteins located in the different clades shown in Fig. 7. The SUOX fold proteins from clades 1 and 2 appear to be associated exclusively with YedZ-like or FdnI-like redox partners. Among these, the majority are members of the YedZ family. Also, there are only seven proteins without a predicted redox partner. Clade 3 shows a large diversity of potential redox partners, with no apparent trend. All of the members of clade 4 contain both redox systems, having an exclusive cytochrome *c* family (family 4) and containing YedZ-like or FdnI-like proteins. Finally, clade 6 also contains an exclusive cytochrome *c* family (family 5). Of the clades predicted to have cytochrome redox partners, clade 6 contains the lowest preponderance of YedZ-like or FdnI-like proteins, where almost 50% of the members do not encode such a protein. Although YedZ homologs or FdnI homologs appear in five of the eight clades of our phylogeny, it is interesting that they are never found in archaea or eukaryotes.

Overall, the presence of SUOX-associated cytochromes correlates with their localizations (Table 4). The SUOX fold proteins predicted to be cytoplasmically located do not appear to have these redox partners and presumably instead rely on cytoplasmic electron donors such as members of the various classes of soluble ferredoxins (59). This is clearly the case with the proteins represented in clades 7 and 8 (Fig. 7). In general, the various classes of cytochromes *c* play an essentially ubiquitous role in periplasmic electron transfer in diderm species and in the mitochondrial intermembrane space. The presence of a YedZ-type or FdnI-type membrane-bound cytochrome *b* correlates with proteins represented in the *tat*-targeted clades shown in Fig. 7. This supports the proposed role of these proteins in coupling SUOX fold-type proteins to the versatile electron sink of the membrane-bound quinone pool. It is also possible that some of the *c*-type cytochromes represented in Table 4 may shuttle electrons between the SUOX fold proteins or complexes and YedZ or FdnI-type membrane-bound cytochromes *b*. These cytochromes *c* may also couple to other membrane-bound redox complexes containing distinct hemes *c*, hemes *b*, and [Fe-S] clusters. Consideration of such systems is beyond the scope of this work (for reviews, see references 2, 71, 74, and 75).

EVOLUTIONARY PATHWAY OF THE SUOX FOLD PROTEINS

Our observation of a large number of SUOX fold proteins across a broad taxonomic diversity presents an opportunity to analyze sequence motif conservation with the aim of presenting an evolutionary pathway encompassing the species from which our sequence data were obtained. As described below, these analyses bear interesting comparisons with the theoretical evolutionary tree proposed by Gupta (27, 28) as well as with evolutionary insights provided by other studies (26, 48, 51, 83).

Motifs

Each of the clades presented in our phylogeny contained at least two relatively conserved motifs (Table 2). In addition, all of the clades, with the exception of clade 3, contain unique sequence motifs. Because of their lack of persistence throughout archeal, bacterial, and eukaryotic species, sequence conservation within the various heme-containing domains or subunits does not provide insights into the SUOX fold class. A similar conclusion can be made about the NADH binding flavoprotein domain of the plant-type NIA enzymes.

In the SUOX fold proteins, the motifs CVExW, CxGNxR, and GxPxR give insight into regions of great importance to the

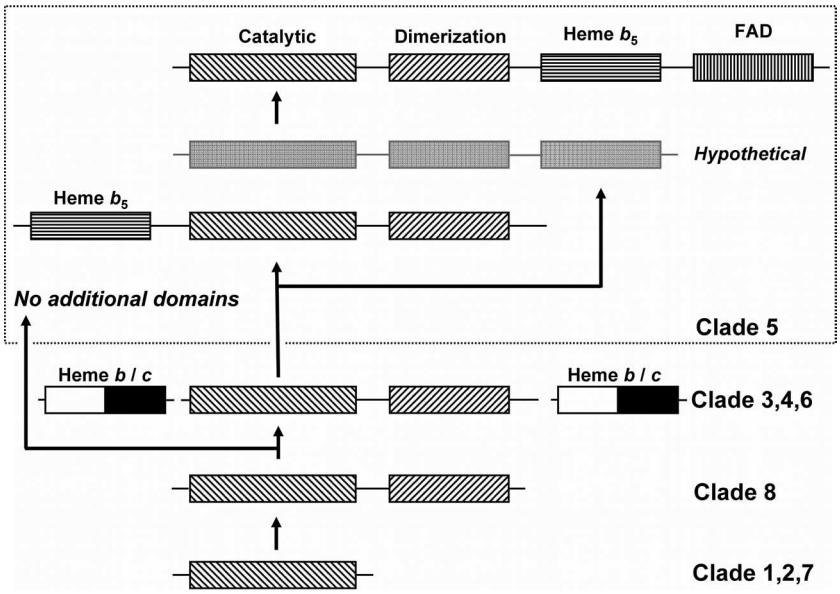

FIG. 13. Potential evolution of the domains found within the SUOX fold protein family. Evolution moves from the simplest form (bottom) to the most complex (top). The dotted outline encompasses a variety of enzymes found in clade 5. The light gray model serves as a hypothetical intermediate between known enzymes. Domains are labeled where appropriate, and lined patterns depict the respective domains in the absence of a label. Clade numbers to the side of the domains indicate the clades where these domain combinations are found. The detached heme *b*/*c* portion indicates a separate gene upstream/downstream of the SUOX fold protein that acts as an interacting subunit.

function and structure of the enzyme. As stated above (see Protein-PPT Interactions), the first two motifs, CVExW and CxGNxR, coordinate the Mo, whereas the GxPxR motif aids in coordinating the phosphate of the PPT. The CVExW and CxGNxR motifs are of great functional importance and are found in seven of the eight clades. The Cys included in these motifs has been identified as being critical for sulfite oxidation because of its role in Mo coordination. These active-site motifs imply two separate evolutionary pathways for the SUOX enzymes, which will be elaborated on below (see "Domains"). The only proteins that do not contain either of these motifs are those found in clade 6, which possess an FxECxxN motif instead. Additional information on active-site evolution can be found in the GxPxR motif. The presence of the GxPxR motif throughout all clades of our phylogeny indicates that while this motif may appear to have a only single interaction with the PPT, the U-shaped turn that it elicits may prove to be structurally important to the active site of this enzyme as well.

The dimerization domain is also found to contain conserved motifs. Of the five structures covered in this review, all contain a dimerization domain, with the exception of YedY. Also, as listed in Table 2, clade 3 does not contain a variety of conserved motifs. Taking this information together, the proteins of known structures represented in clade 5 were prime candidates for analyzing conserved motifs in the dimerization domain. Using this approach, we identified that the RV(E/D)(V/L)S motif was found in the dimerization domain. Along with this conserved motif, there is an abundance of conserved Trp residues in the dimerization domain. We extrapolated this information to clades that do not contain proteins of known structure. We determined the dimerization domain to be an area with a large number of conserved Trp residues found toward the C terminus of the protein. This approach has also identified the GLAWSGxG and SRxxDxTGY motifs, of clade 6 members, as being found in the dimerization domain. In the clades containing proteins without the dimerization domain, conserved motifs exist on the surface of the protein (i.e., NGY in YedY). While YedZ is the proposed redox partner of YedY, this surface motif may allow YedY to interact with other proteins, including additional redox partners.

Domains

Based on the assumption that evolution usually proceeds from less complicated to more complicated architectures, inspection of Fig. 7 and Table 2 suggests that the original SUOX ancestor arose as a monomeric Mo-PPT-containing protein of a species lacking an outer membrane that is represented in clade 7 or clade 8.

The first major event to occur following the emergence of the SUOX fold domain is the addition of a dimerization domain to its C terminus. This represents a fundamental split in the evolution of SUOX fold-containing proteins because, as will be seen below, the SUOX fold domain continues separately through bacterial evolution and is retained as the YedY protein in a large number of proteobacterial species. The first proteins to contain a dimerization domain are in the *Actinobacteria* species represented in clade 3 and in clade 8. It is notable that the actinobacterial and firmicute sequences represented in clade 7 lack this domain. This pathway would ultimately lead to the inclusion of a heme-binding domain as well as the FAD domain (Fig. 13). In the eukaryotic enzymes (clade 5), there appears to be a significant amount of diversity in the presence and sequence of domains. Some eukaryotic organisms, such as fungi and amoebae, contain only the SUOX fold domain and the dimerization domain. The plant SUOX

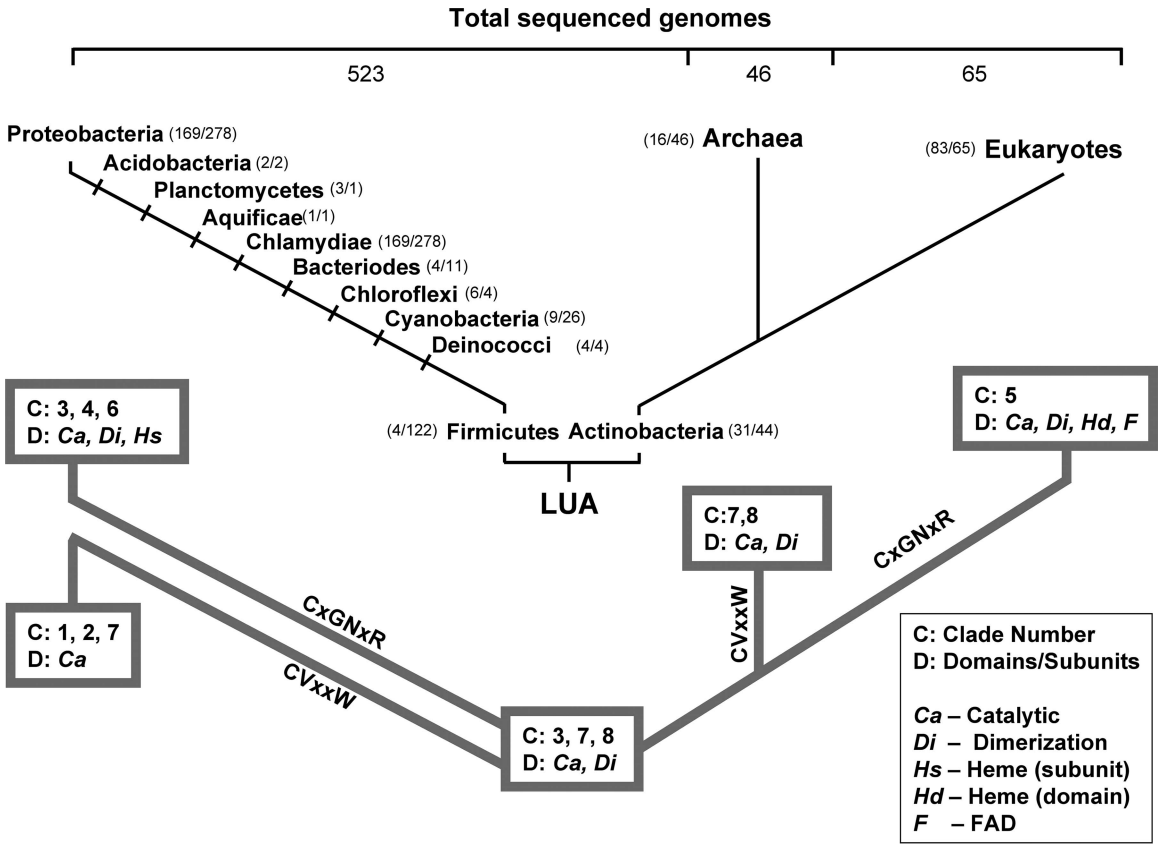

FIG. 14. Evolution of the SUOX family over the evolution of life. The evolutionary path of the SUOX family is overlaid onto a theoretical evolutionary tree of life. Each path is labeled by a conserved active-site motif. The boxes at the end of each branch indicate the clade members (C) that belong to that evolutionary branch as well as the domains present in the SUOX fold proteins (D). Numbers in parentheses listed beside each phylum represent the numbers of genomes solved compared to the numbers of times that the phylum appears in all clades (i.e., number of genomes solved/number of entries in clades).

enzyme shown in Fig. 1 is also included in this group, which finds its place in Fig. 13 by the branch labeled as "no additional domains." The domain shuffling and omissions observed in eukaryotes is not surprising considering that in prokaryotic systems, heme-containing redox partners are found in operons both upstream and downstream of the SUOX protein. As such, a gene fusion of a heme-containing domain to either end is possible.

Following the path of those SUOX fold proteins lacking the dimerization domain, bacterial species that evolved later and are found in clade 7 also synthesize a protein that lacks a dimerization domain. This protein most likely is the progenitor to the proteins found in clades 1 and 2. As clades 1 and 2 are primarily proteobacterial, they evolved later than did half of the species found in clade 7. Additional evidence is found in the motif surrounding the Cys residue that coordinates the Mo in those proteins lacking the dimerization domain. Clade 7 contains the CVTxWS motif that is similar to the CVExW and CVEGW motifs of clades 1 and 2, respectively. It is also notable that while members of clades 1 and 2 are predicted to be exclusively periplasmic by virtue of their *tat* leaders, those of clade 7 are predicted to be cytoplasmically localized.

The evolution of the eukaryotic CxGNxR motif is likely the result of a split occurring in clade 8, where the CVSN Mo-coordinating motif is observed. This clade has the dimerization domain but lacks a *tat* leader. It is likely that the CxGNxR motif evolved from the clade 8 CVSN motif, and it persists in clades 5 (CAGNRR), 4 (CSGNGR), and 3 (CxGNxR), ultimately occurring in the eukaryotic SUOX fold enzymes of clade 5.

Clade 6 appears to have a unique Mo-coordinating motif, FxECxxN, that can be considered to be weakly related to the CAGNRR motif in the sense that there are two residues between the Cys and the Asn. Like clades 1 and 2, this clade is dominated by proteobacterial species, but unlike them, it possesses the dimerization domain.

Correlation with the Evolution of Life

The evolution of the SUOX fold family members can be delineated by taking the appearance of distinct domains and motifs and correlating these with current research regarding the evolution of the three major domains of life (27, 28, 51, 83). Figure 14 shows a potential evolutionary path of the SUOX fold proteins overlaid with a potential evolution of life. In Fig. 14, the two separate evolutionary paths are labeled by the presence of Cys-containing active-site motifs. As the members

of clade 6 show a faint resemblance to the CxGNxR motif, they have been included in this evolutionary branch.

Approximately half of the SUOX fold proteins observed in our study belonged to the *Proteobacteria*. This observation implies two different conclusions. The first is that these systems are heavily weighted toward the *Proteobacteria*. The second is that the completed genomes for bacterial species favors the *Proteobacteria*. If more sequences of proteins for *Proteobacteria* species are known, there is a higher chance of these sequences being included in sequence data sets. We analyzed the Genomes Online Database (http://www.genomesonline.org) and determined that there are far more completed *Proteobacteria* genomes than those of other prokaryotic groups, comprising 277 of the 523 genome sequences. Using our unfiltered list of 559 SUOX fold protein sequences, we compared the number of entries to the number of completed genomes (data not shown). The results show that the apparent favoring of the *Proteobacteria* is created by their larger number of complete genomes. This effect is present in the *Proteobacteria* (168 entries over 278 completed genomes), *Actinobacteria* (31/44), *Bacteriodetes* (4/11), *Chloroflexi* (4/4), and *Aquificae* (1/1), etc. The only exception is the phylum *Firmicutes*, which accounts for only four entries over 122 complete genomes. This information is conveyed in Fig. 14, where the number of completed genomes is found above the respective branches.

Our observation of the relative lack of SUOX fold enzymes in the *Firmicutes* has significant evolutionary implications. Many current theories of evolution place the monoderm bacteria as the first prokaryotes, evolving from the "last universal ancestor" (48). Other theories suggest that diderm cells were the first to evolve, arguing that a bilayered "obcell" was the progenitor of the prokaryotes (9). Our analysis supports the theory that the *Archaea* evolved from monoderm bacteria. In this model, *Firmicutes* species would contain the original ancestor of the SUOX fold enzyme family. This may explain our observation of only two firmicute species being represented in Fig. 7 despite their large number of sequenced genomes (122 competed genomes). The two species are in clade 3 (*Bacillus* sp. strain NRRL B-14911) and clade 7 (*Bacillus licheniformis*). These two entries contain different Cys-containing motifs, where one contains the CxGNxR (clade 3) motif and the other contains the CVTxW motif. In this context, two possibilities exist for SUOX fold evolution. The first is that the *Firmicutes* species contained a single gene that underwent gene duplication and gave rise to two separate SUOX-like gene products. The second is that two separate genes arose in the *Firmicutes*, each containing the SUOX fold (convergent evolution). Either of these possibilities can account for the different Cys-containing motifs. In theory, the archaeal enzymes arose from the CVTxW-containing enzyme of the *Firmicutes*. This CVTxW-containing enzyme would thus be the starting point of the evolutionary branch that contains no additional domains (Fig. 13 and 14). Likewise, the CxGNxR enzyme would be starting point of the evolutionary branch that has the appearance of the dimerization domain. In Fig. 14, we do not suggest that the *Firmicutes* and *Actinobacteria* evolved separately from one another, as may be interpreted by these phyla being at the same evolutionary level. While Gupta (29) suggested that the *Firmicutes* evolved first, a study by Lake et al. (51) suggests that diderms may have evolved from the *Firmicutes* and that the *Archaea* may have evolved from the *Actinobacteria*. Therefore, in Fig. 14, these two phyla have been placed based on the evolution of the SUOX family.

CONCLUSIONS AND OUTLOOK

The explosion of bioinformatic data in recent years has been complemented by an abundance of structural data that enables emerging bioinformatic approaches to be used to gain new insights into SUOX archetypes and closely related proteins. In this review, we used complementary bioinformatic approaches to address the importance and evolutionary persistence of the SUOX fold. These proteins play critical roles in sulfur and nitrogen metabolism and participate in two critical geochemical cycles that are critical for maintaining the integrity and sustainability of the biosphere. We have established that SUOX fold proteins exist in "primordial" archaeal, actinobacterial, or firmicute species and that they persisted throughout evolution to appear in higher animal and plant species.

We have merged structural and sequence bioinformatic data to correlate cladistics analyses with the importance of conserved sequence motifs within eight clades of SUOX fold enzymes. Interestingly, members of many of these clades may have substrates that are not simple inorganic anions such as sulfite or nitrite. It is therefore likely that many SUOX fold enzymes contribute to unanticipated levels of metabolic diversity in a very broad range of bacterial species. Our analyses of sequence conservation should provide interesting targets for site-directed mutagenesis studies that will enable the relationship between protein sequence, Mo-PPT redox chemistry, and substrate specificity to be rigorously explored.

ACKNOWLEDGMENTS

We thank Paul Stothard of the Canadian Bioinformatics Help Desk for authorship of a PERL script that removes closely related entries from within a sequence data set. We thank Craig Knox, also of the Canadian Bioinformatics Help Desk, for writing a PERL script that extracts taxonomies from SWISSPROT and/or TREMBL flat-file entries. We also thank Dean Schieve for installing and maintaining the EMBOSS package used during the course of this work.

This work was funded by the Canadian Institutes of Health Research, the Natural Sciences and Engineering Research Council, and the National Institutes of Health. Infrastructure funding was provided by the Canada Foundation for Innovation. J.H.W. holds a Canada Research Chair in Membrane Biochemistry.

REFERENCES

1. **Berks, B. C.** 1996. A common export pathway for proteins binding complex redox cofactors? Mol. Microbiol. **22:**393–414.
2. **Berks, B. C., S. J. Ferguson, J. W. B. Moir, and D. J. Richardson.** 1995. Enzymes and associated electron transport systems that catalyze the respiratory reduction of nitrogen oxides and oxyanions. Biochim. Biophys. Acta **1232:**97–123.
3. **Berks, B. C., F. Sargent, and T. Palmer.** 2000. The Tat protein export pathway. Mol. Microbiol. **35:**260–274.
4. **Bertero, M. G., R. A. Rothery, N. Boroumand, M. Palak, F. Blasco, N. Ginet, J. H. Weiner, and N. C. Strynadka.** 2005. Structural and biochemical characterization of a quinol binding site of *Escherichia coli* nitrate reductase A. J. Biol. Chem. **280:**14836–14843.
5. **Bertero, M. G., R. A. Rothery, M. Palak, C. Hou, D. Lim, F. Blasco, J. H. Weiner, and N. C. Strynadka.** 2003. Insights into the respiratory electron transfer pathway from the structure of nitrate reductase A. Nat. Struct. Biol. **10:**681–687.
6. **Bones, A. M., and J. T. Rossiter.** 2006. The enzymic and chemically induced decomposition of glucosinolates. Phytochemistry **67:**1053–1067.
7. **Brokx, S. J., R. A. Rothery, G. Zhang, D. P. Ng, and J. H. Weiner.** 2005. Characterization of an *Escherichia coli* sulfite oxidase homologue reveals the

role of a conserved active site cysteine in assembly and function. Biochemistry **44:**10339–10348.
8. **Campbell, W. H.** 2001. Structure and function of eukaryotic NAD(P)H: nitrate reductase. Cell. Mol. Life Sci. **58:**194–204.
9. **Cavalier-Smith, T.** 2001. Obcells as proto-organisms: membrane heredity, lithophosphorylation, and the origins of the genetic code, the first cells, and photosynthesis. J. Mol. Evol. **53:**555–595.
10. **Chan, M. K., S. Mukund, A. Kletzin, M. W. Adams, and D. C. Rees.** 1995. Structure of a hyperthermophilic tungstopterin enzyme, aldehyde ferredoxin oxidoreductase. Science **267:**1463–1469.
11. **Charles, A. M., and I. Suzuki.** 1965. Sulfite oxidase of a facultative autotroph, Thiobacillus novellus. Biochem. Biophys. Res. Commun. **19:**686–690.
12. **D'Errico, G., A. Di Salle, F. La Cara, M. Rossi, and R. Cannio.** 2006. Identification and characterization of a novel bacterial sulfite oxidase with no heme binding domain from *Deinococcus radiodurans*. J. Bacteriol. **188:**694–701.
13. **Di Salle, A., G. D'Errico, F. La Cara, R. Cannio, and M. Rossi.** 2006. A novel thermostable sulfite oxidase from *Thermus thermophilus*: characterization of the enzyme, gene cloning and expression in *Escherichia coli*. Extremophiles **10:**587–598.
14. **Doonan, C. J., U. Kappler, and G. N. George.** 2006. Structure of the active site of sulfite dehydrogenase from *Starkeya novella*. Inorg. Chem. **45:**7488–7492.
15. **Doonan, C. J., H. L. Wilson, K. V. Rajagopalan, R. M. Garrett, B. Bennett, R. C. Prince, and G. N. George.** 2007. Modified active site coordination in a clinical mutant of sulfite oxidase. J. Am. Chem. Soc. **129:**9421–9428.
16. **Douglas, P., N. Morrice, and C. MacKintosh.** 1995. Identification of a regulatory phosphorylation site in the hinge 1 region of nitrate reductase from spinach (*Spinacea oleracea*) leaves. FEBS Lett. **377:**113–117.
17. **Drew, D., D. Sjostrand, J. Nilsson, T. Urbig, C. N. Chin, J. W. de Gier, and G. von Heijne.** 2002. Rapid topology mapping of Escherichia coli inner-membrane proteins by prediction and PhoA/GFP fusion analysis. Proc. Natl. Acad. Sci. USA **99:**2690–2695.
18. **Einsle, O., F. A. Tezcan, S. L. Andrade, B. Schmid, M. Yoshida, J. B. Howard, and D. C. Rees.** 2002. Nitrogenase MoFe-protein at 1.16 Å resolution: a central ligand in the FeMo-cofactor. Science **297:**1696–1700.
19. **Felsenstein, J.** 1992. Estimating effective population size from samples of sequences: a bootstrap Monte Carlo integration method. Genet. Res. **60:**209–220.
20. **Feng, C., G. Tollin, and J. H. Enemark.** 2007. Sulfite oxidizing enzymes. Biochim. Biophys. Acta **1774:**527–539.
21. **Feng, C., H. L. Wilson, J. K. Hurley, J. T. Hazzard, G. Tollin, K. V. Rajagopalan, and J. H. Enemark.** 2003. Essential role of conserved arginine 160 in intramolecular electron transfer in human sulfite oxidase. Biochemistry **42:**12235–12242.
22. **Fischer, K., G. G. Barbier, H. J. Hecht, R. R. Mendel, W. H. Campbell, and G. Schwarz.** 2005. Structural basis of eukaryotic nitrate reduction: crystal structures of the nitrate reductase active site. Plant Cell **17:**1167–1179.
23. **Garrett, R. M., J. L. Johnson, T. N. Graf, A. Feigenbaum, and K. V. Rajagopalan.** 1998. Human sulfite oxidase R160Q: identification of the mutation in a sulfite oxidase-deficient patient and expression and characterization of the mutant enzyme. Proc. Natl. Acad. Sci. USA **95:**6394–6398.
24. **Glaab, J., and W. M. Kaiser.** 1995. Inactivation of nitrate reductase involves NR-protein phosphorylation and subsequent 'binding' of an inhibitor protein. Planta **195:**514–518.
25. **Gonnet, G. H., M. A. Cohen, and S. A. Benner.** 1992. Exhaustive matching of the entire protein sequence database. Science **256:**1443–1445.
26. **Gupta, R. S.** 2001. The branching order and phylogenetic placement of species from completed bacterial genomes, based on conserved indels found in various proteins. Int. Microbiol. **4:**187–202.
27. **Gupta, R. S.** 2005. Molecular sequences and the early history of life, p. 160–183. *In* J. Sapp (ed.), Microbial phylogeny and evolution: concepts and controversies. Oxford University Press, New York, NY.
28. **Gupta, R. S.** 2000. The phylogeny of proteobacteria: relationships to other eubacterial phyla and eukaryotes. FEMS Microbiol. Rev. **24:**367–402.
29. **Gupta, R. S.** 1998. What are archaebacteria: life's third domain or monoderm prokaryotes related to gram-positive bacteria? A new proposal for the classification of prokaryotic organisms. Mol. Microbiol. **29:**695–707.
30. **Hänsch, R., C. Lang, H. Rennenberg, and R. R. Mendel.** 2007. Significance of plant sulfite oxidase. Plant Biol. (Stuttgart) **9:**589–595.
31. **Hänsch, R., C. Lang, E. Riebeseel, R. Lindigkeit, A. Gessler, H. Rennenberg, and R. R. Mendel.** 2006. Plant sulfite oxidase as novel producer of H_2O_2: combination of enzyme catalysis with a subsequent non-enzymatic reaction step. J. Biol. Chem. **281:**6884–6888.
32. **Hänsch, R., and R. R. Mendel.** 2005. Sulfite oxidation in plant peroxisomes. Photosynth. Res. **86:**337–343.
33. **Harris, H. H., G. N. George, and K. V. Rajagopalan.** 2006. High-resolution EXAFS of the active site of human sulfite oxidase: comparison with density functional theory and X-ray crystallographic results. Inorg. Chem. **45:**493–495.
34. **Hille, R.** 2002. Molybdenum and tungsten in biology. Trends Biochem. Sci. **27:**360–367.
35. **Hille, R.** 2002. Molybdenum enzymes containing the pyranopterin cofactor: an overview. Met. Ions Biol. Syst. **39:**187–226.
36. **Hille, R.** 2005. Molybdenum-containing hydroxylases. Arch. Biochem. Biophys. **433:**107–116.
37. **Hille, R.** 1996. The mononuclear molybdenum enzymes. Chem. Rev. **96:**2757–2816.
38. **Johnson, J. L., and M. Duran.** 2001. Molybdenum cofactor deficiency and isolated sulfite oxidase deficiency, p. 3163–3177. *In* C. R. Scryver (ed.), The metabolic and molecular bases of inherited disease. McGraw-Hill, New York, NY.
39. **Jormakka, M., B. Byrne, and S. Iwata.** 2003. Protonmotive force generation by a redox loop mechanism. FEBS Lett. **545:**25–30.
40. **Jormakka, M., S. Tornroth, B. Byrne, and S. Iwata.** 2002. Molecular basis of proton motive force generation: structure of formate dehydrogenase-N. Science **295:**1863–1868.
41. **Kappler, U., and S. Bailey.** 2005. Molecular basis of intramolecular electron transfer in sulfite-oxidizing enzymes is revealed by high resolution structure of a heterodimeric complex of the catalytic molybdopterin subunit and a *c*-type cytochrome subunit. J. Biol. Chem. **280:**24999–25007.
42. **Kappler, U., S. Bailey, C. Feng, M. J. Honeychurch, G. R. Hanson, P. V. Bernhardt, G. Tollin, and J. H. Enemark.** 2006. Kinetic and structural evidence for the importance of Tyr236 for the integrity of the Mo active site in a bacterial sulfite dehydrogenase. Biochemistry **45:**9696–9705.
43. **Kappler, U., B. Bennett, J. Rethmeier, G. Schwarz, R. Deutzmann, A. G. McEwan, and C. Dahl.** 2000. Sulfite:cytochrome *c* oxidoreductase from *Thiobacillus novellus*. Purification, characterization, and molecular biology of a heterodimeric member of the sulfite oxidase family. J. Biol. Chem. **275:**13202–13212.
44. **Karakas, E., H. L. Wilson, T. N. Graf, S. Xiang, S. Jaramillo-Busquets, K. V. Rajagopalan, and C. Kisker.** 2005. Structural insights into sulfite oxidase deficiency. J. Biol. Chem. **280:**33506–33515.
45. **Kay, C. J., M. J. Barber, B. A. Notton, and L. P. Solomonson.** 1989. Oxidation—reduction midpoint potentials of the flavin, haem and Mo-pterin centres in spinach (*Spinacia oleracea* L.) nitrate reductase. Biochem. J. **263:**285–287.
46. **Kisker, C., H. Schindelin, D. Baas, J. Retey, R. U. Meckenstock, and P. M. Kroneck.** 1998. A structural comparison of molybdenum cofactor-containing enzymes. FEMS Microbiol. Rev. **22:**503–521.
47. **Kisker, C., H. Schindelin, A. Pacheco, W. A. Wehbi, R. M. Garrett, K. V. Rajagopalan, J. H. Enemark, and D. C. Rees.** 1997. Molecular basis of sulfite oxidase deficiency from the structure of sulfite oxidase. Cell **91:**973–983.
48. **Koch, A. L.** 2003. Were gram-positive rods the first bacteria? Trends Microbiol. **11:**166–170.
49. **Krissinel, E., and K. Henrick.** 2004. Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. Acta Crystallogr. D Biol. Crystallogr. **60:**2256–2268.
50. **Krogh, A., B. Larsson, G. von Heijne, and E. L. Sonnhammer.** 2001. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J. Mol. Biol. **305:**567–580.
51. **Lake, J. A., C. W. Herbold, M. C. Rivera, J. A. Servin, and R. G. Skophammer.** 2007. Rooting the tree of life using nonubiquitous genes. Mol. Biol. Evol. **24:**130–136.
52. **Lopez-Huertas, E., F. J. Corpas, L. M. Sandalio, and L. A. Del Rio.** 1999. Characterization of membrane polypeptides from pea leaf peroxisomes involved in superoxide radical generation. Biochem. J. **337:**531–536.
53. **Loschi, L., S. J. Brokx, T. L. Hills, G. Zhang, M. G. Bertero, A. L. Lovering, J. H. Weiner, and N. C. Strynadka.** 2004. Structural and biochemical identification of a novel bacterial oxidoreductase. J. Biol. Chem. **279:**50391–50400.
54. **Lu, G., Y. Lindqvist, G. Schneider, U. Dwivedi, and W. Campbell.** 1995. Structural studies on corn nitrate reductase: refined structure of the cytochrome *b* reductase fragment at 2.5 Å, its ADP complex and an active-site mutant and modeling of the cytochrome *b* domain. J. Mol. Biol. **248:**931–948.
55. **MacKintosh, C., and S. E. Meek.** 2001. Regulation of plant NR activity by reversible phosphorylation, 14-3-3 proteins and proteolysis. Cell. Mol. Life Sci. **58:**205–214.
56. **Marquez, A. J., M. Betti, M. Garcia-Calderon, P. Pal'ove-Balang, P. Diaz, and J. Monza.** 2005. Nitrate assimilation in *Lotus japonicus*. J. Exp. Bot. **56:**1741–1749.
57. **Mendel, R. R.** 2007. Biology of the molybdenum cofactor. J. Exp. Bot. **58:**2289–2296.
58. **Mendel, R. R., and F. Bittner.** 2006. Cell biology of molybdenum. Biochim. Biophys. Acta **1763:**621–635.
59. **Meyer, J.** 2001. Ferredoxins of the third kind. FEBS Lett. **509:**1–5.
60. **Moorhead, G., P. Douglas, N. Morrice, M. Scarabel, A. Aitken, and C. MacKintosh.** 1996. Phosphorylated nitrate reductase from spinach leaves is inhibited by 14-3-3 proteins and activated by fusicoccin. Curr. Biol. **6:**1104–1113.
61. **Moser, C. C., T. A. Farid, S. E. Chobot, and P. L. Dutton.** 2006. Electron tunneling chains of mitochondria. Biochim. Biophys. Acta **1757:**1096–1109.

62. **Moser, C. C., J. M. Keske, K. Warncke, R. S. Farid, and P. L. Dutton.** 1992. Nature of biological electron transfer. Nature **355:**796–802.
63. **Moser, C. C., C. C. Page, and P. L. Dutton.** 2006. Darwin at the molecular scale: selection and variance in electron tunnelling proteins including cytochrome *c* oxidase. Philos. Trans. R. Soc. Lond. B Biol. Sci. **361:**1295–1305.
64. **Moser, C. C., C. C. Page, R. Farid, and P. L. Dutton.** 1995. Biological electron transfer. J. Bioenerg. Biomem. **27:**263–274.
65. **Mudd, S. H., F. Irreverre, and L. Laster.** 1967. Sulfite oxidase deficiency in man: demonstration of the enzymatic defect. Science **156:**1599–1602.
66. **Mukund, S., and M. W. W. Adams.** 1990. Characterization of a tungsten-iron-sulfur protein exhibiting novel spectroscopic and redox properties from the hypothermophilic archaebacterium *Pyrococcus furiosus*. J. Biol. Chem. **265:**11508–11516.
67. **Pacheco, A., J. T. Hazzard, G. Tollin, and J. H. Enemark.** 1999. The pH dependence of intramolecular electron transfer rates in sulfite oxidase at high and low anion concentrations. J. Biol. Inorg. Chem. **4:**390–401.
68. **Page, C. C., C. C. Moser, and P. L. Dutton.** 2003. Mechanism for electron transfer within and between proteins. Curr. Opin. Chem. Biol. **7:**551–556.
69. **Peters, J. W., K. Fisher, and D. R. Dean.** 1995. Nitrogenase structure and function: a biochemical-genetic perspective. Annu. Rev. Microbiol. **49:**335–366.
70. **Rice, P., I. Longden, and A. Bleasby.** 2000. EMBOSS: the European Molecular Biology Open Software Suite. Trends Genet. **16:**276–277.
71. **Richardson, D. J.** 2000. Bacterial respiration: a flexible process for a changing environment. Microbiology **146:**551–571.
72. **Richardson, D. J.** 2001. Introduction: nitrate reduction and the nitrogen cycle. Cell. Mol. Life Sci. **58:**163–164.
73. **Richardson, D. J., B. C. Berks, D. A. Russell, S. Spiro, and C. J. Taylor.** 2001. Functional, biochemical and genetic diversity of prokaryotic nitrate reductases. Cell. Mol. Life Sci. **58:**165–178.
74. **Richardson, D. J., and N. J. Watmough.** 1999. Inorganic nitrogen metabolism in bacteria. Curr. Opin. Chem. Biol. **3:**207–219.
75. **Rothery, R. A., G. J. Workun, and J. H. Weiner.** 2007. The prokaryotic complex iron-sulfur molybdoenzyme family. Biochim. Biophys. Acta. [Epub ahead of print.] doi:10.1016/j.bbamem.2007.09.002.
76. **Saitou, N., and M. Nei.** 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. **4:**406–425.
77. **Sanchez-Pulido, L., A. M. Rojas, A. Valencia, A. C. Martinez, and M. A. Andrade.** 2004. ACRATA: a novel electron transfer domain associated to apoptosis and cancer. BMC Cancer **4:**98.
78. **Sargent, F.** 2007. Constructing the wonders of the bacterial world: biosynthesis of complex enzymes. Microbiology **153:**633–651.
79. **Sargent, F., B. C. Berks, and T. Palmer.** 2002. Assembly of membrane-bound respiratory complexes by the Tat protein-transport system. Arch. Microbiol. **178:**77–84.
80. **Schindelin, H., C. Kisker, J. Hilton, K. V. Rajagopalan, and D. C. Rees.** 1996. Crystal structure of DMSO reductase: redox-linked changes in molybdopterin coordination. Science **272:**1615–1621.
81. **Schneider, F., J. Löwe, R. Huber, H. Schindelin, C. Kisker, and J. Knäben.** 1996. Crystal structure of dimethyl sulfoxide reductase from *Rhodobacter capsulatus* at 1.88 Å resolution. J. Mol. Biol. **263:**53–69.
82. **Schrader, N., K. Fischer, K. Theis, R. R. Mendel, G. Schwarz, and C. Kisker.** 2003. The crystal structure of plant sulfite oxidase provides insights into sulfite oxidation in plants and animals. Structure **11:**1251–1263.
83. **Skophammer, R. G., C. W. Herbold, M. C. Rivera, J. A. Servin, and J. A. Lake.** 2006. Evidence that the root of the tree of life is not within the Archaea. Mol. Biol. Evol. **23:**1648–1651.
84. **Solomonson, L. P., M. J. Barber, W. D. Howard, J. L. Johnson, and K. V. Rajagopalan.** 1984. Electron paramagnetic resonance studies on the molybdenum center of assimilatory NADH:nitrate reductase from *Chlorella vulgaris*. J. Biol. Chem. **259:**849–853.
85. **Sonnhammer, E. L. L., G. von Heijne, and A. Krogh.** 1998. A hidden Markov model for predicting transmembrane helices in protein sequences. Proc. Int. Intell. Syst. Mol. Biol. **6:**175–182.
86. **Spence, J. T., C. A. Kipke, J. H. Enemark, and R. A. Sunde.** 1991. Stoichiometry of electron uptake and the effect of anions and pH on the molybdenum and heme reduction potentials of sulfite oxidase. Inorg. Chem. **30:**3011–3015.
87. **Thompson, J. D., T. J. Gibson, F. Plewniak, F. Jeanmougin, and D. G. Higgins.** 1997. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. **25:**4876–4882.
88. **Thompson, J. D., D. G. Higgins, and T. J. Gibson.** 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. **22:**4673–4680.
89. **von Rozycki, T., M. R. Yen, E. E. Lende, and M. H. Saier, Jr.** 2004. The YedZ family: possible heme binding proteins that can be fused to transporters and electron carriers. J. Mol. Microbiol. Biotechnol. **8:**129–140.
90. **Weiner, J. H., P. T. Bilous, G. M. Shaw, S. P. Lubitz, L. Frost, G. H. Thomas, J. A. Cole, and R. J. Turner.** 1998. A novel and ubiquitous system for membrane targeting and secretion of cofactor-containing proteins. Cell **93:**93–101.
91. **Wu, L.-F., B. Ize, A. Chanal, Y. Quentin, and G. Fichant.** 2000. Bacterial twin-arginine signal peptide-dependent protein translocation pathway: evolution and mechanism. J. Mol. Microbiol. Biotechnol. **2:**179–189.
92. **Yamanaka, T., T. Yoshioka, and K. Kimura.** 1981. Purification of sulphite-cytochrome *c* reductase of *Thiobacillus novellus* and reconstitution of its sulphite oxidase system with the purified constituents. Plant Cell Physiol. **22:**613–622.
93. **Yankovskaya, V., R. Horsefield, S. Törnroth, C. Luna-Chavez, H. Miyoshi, C. Léger, B. Byrne, G. Cecchini, and S. Iwata.** 2003. Architecture of succinate dehydrogenase and reactive oxygen species generation. Science **299:**700–704.
94. **Zhang, Z., L. Huang, V. Shulmeister, Y. Chi, K. K. Kim, L. Hung, A. R. Crofts, E. A. Berry, and S. Kim.** 1998. Electron transfer by domain movement in cytochrome bc_1. Nature **392:**677–684.